THE WORLD BANK

RESEARCH OBSERVER

The World Bank Research Observer is intended for anyone who has a professional interest in development. *Observer* articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the *Observer* considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the editor. Please direct all editorial correspondence to the Editor, *The World Bank Research Observer*, at the address in the copyright notice below.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of the data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

The World Bank Research Observer is published twice a year in February and August. Single copies may be purchased for US$12.95. Subscription rates are as follows:

	Individuals	Institutions
1-year subscription	US$25	US$40
2-year subscription	US$45	US$[illegible]5
3-year subscription	US$60	US$100

Orders should be sent to: World Bank Publications, Box 7247-7956, Philadelphia, Pennsylvania 19170-7956, U.S.A. Readers with mailing addresses in developing countries and in socialist economies in transition are eligible to receive complementary subscriptions on request. Written confirmation is required every two years to renew such subscriptions.

ISSN 0257-3032

The World Bank Research Observer is indexed by the *Journal of Economic Literature*, the *Standard Periodical Directory*, ABI/INFORM, the Index to International Statistics, the Public Affairs Information Service, Inc., the Social Sciences Citation Index, and, online, by the Economic Literature Index and DIALOG. It is available in microform through University Microfilms, Inc., 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

THE WORLD BANK

RESEARCH OBSERVER

VOLUME 11 | NUMBER 2 | AUGUST 1996

THE WORLD BANK

Research Observer

Volume 12 • Number 1 • February 1997

THE WORLD BANK

Research Observer

Volume 12 • Number 2 • August 1997

THE WORLD BANK

Research Observer

Volume 13 • Number 1 • February 1998

THE WORLD BANK

Research Observer

Volume 13 • Number 2 • August 1998

THE WORLD BANK

Research Observer

Volume 14 • Number 1 • February 1999

SOME LESSONS FROM THE EAST ASIAN MIRACLE

Joseph E. Stiglitz

The rapid economic growth of eight East Asian economies, often called the "East Asian miracle," raises two questions: What policies and other factors contributed to that growth? And can other developing countries replicate those policies to stimulate equally rapid growth?

This article, based on case studies, econometric data, and economic theory, offers a list of the ingredients that contributed to that success. But it is the combination of these ingredients, many of which involve government interventions acting together, that accounts for East Asia's success.

The remarkable success of the economies of East Asia raises the question: to what can that success be attributed? In most of the eight economies that are part of the "East Asian miracle"—Hong Kong, Indonesia, Japan, the Republic of Korea, Malaysia, Singapore, Taiwan (China), and Thailand—government undertook major responsibility for the promotion of economic growth. Which policies contributed to the success of these economies, and why? Ascertaining what would have happened in the absence of the specified policy is often difficult. That the government subsidized a sector that grew rapidly does not imply that the growth should be attributed to the government's action. The sector might have grown without government intervention.

This article is an interpretive essay based on case studies, econometric data, and economic theory. In formulating a coherent explanation of East Asia's experience, I do not present a formula or a simple recipe, but rather a list of ingredients. Because these ingredients are interactive, and because they were introduced in conjunction with other policies, the government's approach has to be evaluated as a package. Indeed, East Asia's success was based on a combination of factors, particularly the high savings rate interacting with high levels of human capital accumulation, in a stable, market oriented environment—but one with active government intervention—that was conducive to the transfer of technology.[1]

Each of the economies is unique; each differs in its history and culture. Some, such as Singapore and Hong Kong, are small city-states. Others are large. Many

are racially and culturally homogeneous; some, such as Malaysia, are culturally diverse. But it seems implausible to attribute the success of each of these countries to special factors; the task instead is to discover the common threads.

Moreover, the unique factors typically refer to certain cultural aspects, such as a Confucian heritage, that are suspect: not that long ago, the Confucian heritage, with its emphasis on traditional values, was cited as an explanation for why these countries had not grown. To be sure, cultural factors may play an important role: the stress on education has contributed much to the success of these countries.

Statistical Explanations

Having expressed reservations about the usefulness of cultural explanations, it is important to note as well the limitations to the standard statistical techniques. For almost four decades (Solow 1957), the standard approach has been to ask to what extent this growth can be explained by increases in inputs, that is, human and physical capital, and expenditures for the acquisition of technology. In this approach, the "miracle" is the amount of this growth that cannot be so explained (the residual). Several such studies have argued that the East Asian experience can largely be explained by rapid increases in inputs—high levels of investment and heavy expenditures for education. Krugman (1994) and Young (1993, 1995) argue that essentially all of the growth in Singapore can be so explained. Others, such as Kim and Lau (1993) and World Bank (1993), find more evidence of a positive residual, but even then it is not unusually high. There are a variety of technical reasons why the applicability of this methodology to at least several of the East Asian countries should be suspect. Whatever their flaws, however, these studies still offer an important lesson: policies that increase the accumulation of physical and human capital are likely to lead to more rapid growth. The real problem is what these studies leave unanswered. The unique and changing circumstances of each country and the multitude of programs involved, each with a number of potentially important features, imply that a statistical study would be relevant only for addressing the broadest questions: were savings rates unusually high, or were financial restraints associated with faster rates of economic growth? Such studies do not identify those features that facilitated what in retrospect was a remarkable transformation of the economy. To understand this transformation, answers must be found to the following:

First, why were saving rates so high? Elsewhere, such saving rates had only been attained under the compulsion of strong government force, as in the Communist countries. Although studies suggest that these saving rates may be explained in part by economic factors—such as high growth rates that spurred high saving rates, as consumption lagged increases in incomes (Carroll, Weil, and Summers 1993; Stiglitz and Uy 1996)—government actions also played an

important role in mobilizing savings (although this "virtuous cycle" between growth and savings was important as well).

Second, how was it possible to invest efficiently at such a rapid pace? To be sure, if life consisted of nothing more than adding homogeneous capital to a homogeneous production process, East Asia's success would hardly be remarkable. But in that case other countries that attempted to invest rapidly would have had far more success than they have had.

Third, how was it possible to reduce the technology gap so quickly? Clearly, more was entailed than just buying technology. To encourage the transfer of technology from foreign investors, the East Asian economies made enormous investments in human capital, educating large numbers of skilled engineers able to absorb and adapt the most advanced technology. And the East Asian economies were willing to accept foreign investment and create an economic atmosphere conducive to its entry.[2] Moreover, they combined these efforts with an emphasis on the most technologically advanced investment.

And finally how did East Asia ensure that the benefits of rapid growth were spread widely among the population? Previous theories suggested that rapid growth was associated with rapid capital accumulation, which in turn was associated with high degrees of inequality; and that growth would in fact be accompanied by an increase in the degree of inequality (Kuznets 1955). Not only did this assumption prove to be false, but there are reasons to believe that government policies that promoted greater equality contributed in no small measure to the remarkable growth of these countries.

Metaphors of Economic Growth

Several metaphors are used to describe the process of economic growth. These metaphors undoubtedly influence how we think about the subject.

An Engine Metaphor

Perhaps the most popular metaphor is that which refers to the engine of growth—as if there were a motor driving the performance of the economy. Capital accumulation is often given credit for being the engine of growth. And the countries of East Asia certainly have accumulated capital at an impressive rate. Sometimes the concept of capital accumulation is broadened to include human capital—the improvement in the skills of the labor force. And sometimes these two are given credit not only for their direct contribution, but also for the technical progress that might not have occurred in their absence.

Once one identifies the engine of growth, one tries to make the engine stronger. Thus, if capital accumulation is the engine, the task is to increase capital accumulation. The role of government is to rev up the engine to encourage a higher rate of capital accumulation. The engine metaphor has some important

limitations: it encourages a search for particular factors that account for growth, although it may in fact be the system as a whole, including the interactions among the parts, that accounts for growth. If human capital accumulation is inadequate, even rapid physical capital accumulation may be ineffective. But if both are required, which one is *the* engine?

A Chemical Metaphor

I prefer two other metaphors for thinking about the growth process. One is borrowed from chemistry: the government as catalyst. The government can be a catalyst for growth without necessarily providing a great deal of resources. Indeed, that is the remarkable property of catalysts—having set off a chemical reaction, they are themselves not used up in the process. At the very least this metaphor warns that the effect of some government policy should not be measured simply by asking about the magnitude of the subsidy or what fraction of the funds was provided by the government. More concretely, investments in human capital and infrastructure, both physical and institutional, can increase the private return to investment and thereby promote growth.

A Biological Metaphor

The second metaphor, "adaptive systems," is borrowed from biological terminology. Species that survive adapt to changes in their surroundings. More advanced species survive in part because of their ability to learn. Thus the most important characteristic for survival is not a particular policy, but the ability to respond to changes in the environment and to learn from past mistakes. Adaptability is often said to distinguish private sector enterprises from government bureaucracy. But government, because of its monopoly powers, can survive even if it does not adapt well or quickly. The East Asian economies demonstrated that government too can be highly adaptive. When changes in the environment made previously adopted policies inappropriate, these governments changed course, and they learned quickly from their mistakes. As their economies grew and became more complex, the state's role clearly had to change. there was neither the need nor the capacity for active intervention on the scale previously assumed. And officials recognized the importance of adopting policies to promote higher levels of technology and higher value-added industries.

A Metaphor from Physics

One metaphor has been omitted from this list—the one that has in fact dominated the economics profession for almost a century—the economy as an equilibrium system. The omission is deliberate: it is a metaphor that provides little insight into the dramatic changes that occurred in these societies. The equilibrium metaphor suggests that individuals had (perhaps rational) expectations

concerning future rates of return; given those expectations, they determined their saving rate; meanwhile, profit-maximizing firms scoured the world looking for the best products and technologies to employ, given the costs of adjustment. In this metaphor, too, government played at most an ancillary role. This metaphor leaves unasked—and unanswered—such fundamental questions as, What set the East Asian countries apart from other countries? Why is their experience so different?

Complementing Markets Rather Than Replacing Them

Before the East Asia miracle there were two dominant paradigms for development, one focused on markets, the other on government and planning. The first had its intellectual roots in Adam Smith's "invisible hand": markets lead to efficient outcomes. All that government needs to do to promote growth is get out of the way. The basic slogan is "get the prices right." With the right prices, everyone will have an incentive to make the right resource allocations. Undermining this particular religion was the disturbing observation that countries that seemed to get the prices right—to follow all the advice of the visiting preachers of the free market—too often failed to grow. To be sure, like medieval medicine, there was always the allegation that the patient had not followed the doctor's orders precisely, and it was this that accounted for the failure of the remedy.

At the opposite side were those who had little faith in the market and who looked to government to ensure through the planning process that resources were deployed in a way that promoted economic growth. The lack of success of those countries that followed this paradigm has led to the virtual extinction of this school of thought.

Ironically, almost none of the successful industrial countries followed either of these extreme strategies. They are mixed economies in which government plays an important role. The appropriate question to be asked is not whether government should play a role, but what role and how can it be performed most effectively.

At the same time that the success of the East Asian economies and the collapse of the socialist economies called into question the standard paradigm, advances in economic theory called into question the intellectual foundations of these two approaches. In the mid-1950s Arrow and Debreu (1954) identified several conditions that must be satisfied if markets are to yield efficient outcomes. These include, first, the absence of externalities (external economies or diseconomies that affect the activity in question) and of public goods (commodities or services that, once provided, can be obtained without payment by others); second, the presence of perfect competition; and, third, a complete set of markets, including markets extending infinitely far into the future and covering all risks. A market failure is said to occur where these conditions are not satisfied. This approach identified specific interventions by the government to

correct each market failure, for instance, pollution taxes to correct for environmental damage. Government had a well-defined, highly circumscribed role.

It was not until thirty years later, however, that the full limits of the market mechanism became well understood. Hidden in Arrow and Debreu's framework were strong assumptions about information and technology. In their model information need not be perfect, but it could not change as a result of actions taken within the economy. Greenwald and Stiglitz (1986) showed that whenever information was imperfect or markets were incomplete, government could devise interventions that filled in for these imperfections and that could make everyone better off. Because information was never perfect and markets never complete, these results completely undermined the standard theoretical basis for relying on the market mechanism. Similarly the standard models ignored changes in technology; for a variety of reasons markets may underinvest in research and development (see, for example, Stiglitz 1987, 1988, and Arrow 1962). Because developing economies have underdeveloped (missing) markets and imperfect information and because the development process is associated with acquiring new technology (new information), these reservations about the adequacy of market mechanisms may be particularly relevant to developing countries (Stiglitz 1989).

The modern theory of market failures recognizes, however, that government interventions may not actually improve matters. Theories of regulatory capture and rent-seeking imply that government interventions may contribute to inefficient resource allocation, and whatever their weaknesses, these theories have sufficient plausibility to suggest that governments need to exercise caution. How the government intervenes may matter a great deal.

The fundamental mistake of the countries of the former Soviet Union and those developing countries that tried to rely on planning was that they sought to correct market failures by replacing the market. The governments of East Asia, by contrast, recognized the limitations of markets but confined the government's role to

- Policies that actively sought to ensure macroeconomic stability.
- Making markets work more effectively by, for instance, regulating financial markets.
- Creating markets where they did not exist.
- Helping to direct investment to ensure that resources were deployed in ways that would enhance economic growth and stability.
- Creating an atmosphere conducive to private investment and ensured political stability.

In short, rather than replacing markets, these governments promoted and used them. Such interventions had to be carefully balanced; if they were too heavy-handed, they might have squelched the market. This agenda required government to design interventions in a way that reduced the likelihood of rent-seeking behavior and that increased its ability to adapt to changing circum-

stances. One such mechanism was a performance-based reward structure that provided strong growth-oriented incentives and served as a basis for awarding government subsidies. This structure was relatively free from corruption and helped to direct resources to areas that produced high economic returns. Another essential step was to design a civil service system based on merit, which compensated employees well and built in provisions that reduced the dangers of corruption.

In this discussion, the interventions are organized around four major themes: industrial policies, cooperation and competition, equality, and export-led growth. Some of the most important actions to promote economic growth were directed to the financial market, and these interventions are the subject of the accompanying article in this journal by Stiglitz and Uy.

Industrial Policies

Industrial policies are directed at developing and encouraging certain sectors. What were these industrial policies? Why were they adopted? And did they work, either by directing resources to desired areas or, more broadly, in promoting economic growth?

What Policies Were Pursued?

Most countries shared three objectives: developing technological capabilities, promoting exports, and building the domestic capacity to manufacture a range of intermediate goods (such as plastics and steel). Support for particular industries and imports of the necessary foreign technology took several forms. First, the support for education—particularly engineering and science education—provided an intellectual infrastructure that facilitated technological transfer. Second, the decision to discourage (through financial market regulations) the allocation of capital to areas such as real estate meant that more capital was available for areas with higher technological benefits, such as plants and equipment. Third, as discussed later, the government encouraged exports. Fourth, in some industries, particularly those with many firms, government promoted technology programs, including science centers that offered services ranging from identifying new products to providing research and development for firms that had no facilities of their own. Taiwan (China) and Malaysia developed industrial parks for high-technology industries, both to allow firms to capture some of the diffuse externalities associated with these industries as well as to lower the barriers to entry. (Diffuse externalities arise when the actions of one firm benefit—or confer costs upon—many firms, rather than, say, just one upstream firm or one downstream firm.) And finally, the government provided explicit and implicit subsidies (through cheap credit) to industries it wished to support.

An important element in the expansion of certain industries was a receptivity to direct foreign investment. The East Asian economies not only resisted xenophobic aversions to foreign investments, but they also induced capital inflows by providing sound macroeconomic management, a stable political environment, and well-managed labor markets with educated workers. In many cases governments took explicit steps to ensure that a transfer of technological and human capital would accompany these inflows. Foreign investment increased the pace of expansion, reducing the constraints imposed by limitations on the availability of capital, domestic entrepreneurship, and technological know-how.

Why Were Industrial Policies Adopted?

Market failures are likely to be particularly significant in developing countries for several reasons.[3] Understanding these market failures helps explain the policies that were adopted and the reasons they were so effective.

WEAK AND NONEXISTENT MARKETS. In the early stages of development, markets often do not exist or work well, so prices may not provide good signals for resource allocation. In East Asia capital markets were particularly weak, leading government to create institutions to promote savings (the postal savings banks) and to extend long-term credit (the development banks). Governments also tried to develop the financial infrastructure by helping to establish bond and equity markets (Stiglitz and Uy 1996).

Having promoted savings, governments had to decide how to allocate these funds. If there had been well-established market institutions for allocating long-term capital, governments could have made use of those institutions. But because the governments had to decide how to allocate resources, it was natural to direct the funds to projects that would yield the highest level of social welfare.

TECHNOLOGICAL SPILLOVERS. Private markets have inadequate incentives for investing in the production and acquisition of technology, largely because it is difficult to appropriate the returns to knowledge. Developing countries typically operate at a level of technology far below that of industrial countries; development is, to a large extent, the process of acquiring and adapting existing technologies. Patent protection ensures that the seller can command some payment for new technology, but it does not provide much protection for a firm that transfers and adapts an existing technology. Adopting and adapting new technologies involves a risk. If successes are quickly imitated, then firms face a "heads I lose, tails you win" situation: when they succeed, there is little profit because of the force of competition; when they fail, they lose money.

MARKETING SPILLOVERS. Still another kind of valuable information concerns marketing. Knowing where there is a market for a product is not information that can be kept secret. If a firm spends money to discover that Americans like

madras shirts, then any manufacturer of madras shirts can take advantage of that information. The converse is that the products of a country establish a reputation. Thus, Japan's reputation for high quality benefits all Japanese producers.

Such marketing spillovers have led governments to adopt programs aimed at promoting the country's products. (In Hong Kong these programs are financed by a special tax. In Singapore they are directed by the powerful Economic Development Board.) Spillovers have also resulted in an array of programs to improve the countries' reputation. Most notable in this respect is the recent effort by Taiwan (China) to encourage its domestic firms to obtain brand recognition.

RETURNS TO SCALE: a problematic explanation. Not all of the arguments advanced as rationales for industrial policies are persuasive, however. One that seemed particularly influential in Japan held that government intervention was required to rationalize industry. It was argued that without government support, firms would be too small, and the large number of such firms would reduce the profitability of all firms in a sector. (Thus, the Japanese government not only condoned the increased concentration in the steel industry in the late 1960s but, in one of its most famous mistakes, tried to discourage Honda—at the time a successful manufacturer of motorcycles—from entering the automobile market.) This argument is unpersuasive because if there truly were increasing returns to scale, then a single firm would benefit by increasing its production; in time its costs would be lowered, and it would then be able to undercut its rivals. Natural economic forces lead to the rationalization of industries without government intervention.

A slight variant of the argument about returns to scale does have some validity. Increasing returns combined with a shortage of capital may stunt small firms. They cannot expand to take advantage of increasing returns either because they cannot get access to capital or because the only form of capital to which they have access is credit, which imposes too high a risk. In this case, government intervention can lower the costs of capital and increase economic efficiency.

Increasing returns, especially when combined with capital market imperfections, provide the foundation for strategic trade policy. Historically, arguments for government trade interventions focused on industries with learning by doing. If today's production lowers future marginal costs, that creates a form of increasing returns akin to the more familiar static increasing returns. A firm that expands production lowers its future production costs and undercuts its rivals. The infant industry argument holds that protection is important so that the young firm can gain the experience required to lower its production costs and allow it to become viable. Critics of this argument claim that if the firm is to be profitable in the long run, it should incur any necessary losses today. But this assumption is based on the premise that capital markets are perfect. With imperfect capital markets, a firm may not be able to sustain the losses that would enable it to produce at a level at which it would eventually become profitable.

Moreover, if the firm is unable to appropriate all the returns to its learning, then social returns to production will exceed private returns (Dasgupta and Stiglitz 1988). In addition, dominant firms in industrial countries are likely to take advantage of the lack of competition that prevails when learning is important by raising prices and increasing their profits. Government policies may be directed at trying to appropriate some of these rents (the excess profits that result from a dominant competitive position).

COORDINATION FAILURES. The widespread absence of markets in developing countries means that prices cannot perform their coordination role. Government may thus have to assume a more active role in performing this function. The traditional examples relate to the development of downstream and upstream industries: developing a steel-manufacturing industry does not pay unless there is a steel-using industry; and developing a steel-using industry does not pay if there is no steel-manufacturing industry. If both wait, nothing happens. According to this view, the government has an important function in coordinating the two activities. Such coordination failures, it is argued, are likely to be most important when the returns to scale are large. For instance, if manufacturing steel is deemed to be desirable, it is necessary to build a large steel plant and a large steel-using industry. Other market failures, such as the absence of risk markets, interact with this failure: large risks are likely to accompany such large scale investments, and the market provides no mechanism by which these risks can be divested. Moreover, no single entrepreneur could amass the capital required, and the imperfections of the capital market mean that it cannot supply the funds required. Developing countries are less likely than industrial countries to have the organizations capable of undertaking these large investments in a single sector, let alone the capacity to undertake the investments in both the upstream and downstream firms. Thus coordination problems may be larger in developing countries, and the capacity to deal with them may be smaller.

The earlier arguments for coordination failures (Rosenstein Rodan 1943 and Murphy, Shleifer, and Vishny 1989) were rightly criticized as unpersuasive (Stiglitz 1994a). Such a problem could easily be addressed through trade—one of the solutions devised by the East Asian countries (without benefit of the theoretical literature). It is possible to develop a steel-using industry simply by importing steel and to develop steel producers without steel users simply by exporting steel.

In the early stages of rapid growth, the subsectors responsible for the takeoff in many, if not most, of the East Asian countries—textiles, footwear, sporting goods, toys—were not those in which economies of scale or coordination problems seemed important. But there was a more subtle form of returns to scale in which government intervention did matter and which affected growth even in these areas: the availability of a wide range of intermediate—often fairly complex—goods, tailored for the producers of final goods. The sellers of these intermediate goods do not capture all of the benefits that their greater availability

provides. The improved two-way flow of information between the producer and the user, which permits better coordination in the development of the intermediate and final goods, is a benefit of proximity. That explains why importing the intermediate good does not serve as a perfect substitute for domestic production and also provides a rationale for government intervention. In Malaysia it is claimed that the local auto manufacturer has provided important spillovers to the intermediate goods firms that produce parts and that these firms, in turn, have benefited producers of other final goods.

STRATEGIC NEGOTIATIONS. In negotiations with other countries or companies, the governments of East Asia have often recognized—and taken advantage of—the nature of the market environment. The outcome of any bargaining depends on the strength of competition on both sides. By reducing competition among buyers of technology and trying to increase competition among sellers, the governments succeeded in appropriating more of the surplus associated with the transfer of technology than otherwise could have been captured. In Japan, for instance, a single firm was sometimes given the right to negotiate a licensing agreement; it might then be compelled to share the technology with other firms in the industry.

Did These Policies Work?

Industrial policies have been widely criticized, on the (somewhat contradictory) grounds that they were ineffective or distortionary. The first criticism suggests that industrial policies are more form than substance. Critics cite statistics such as the small percentage of loans made by the development banks. These statistics are unconvincing, however: the consequences of, say, a loan by the Industrial Bank of Japan may be far greater than the actual dollars lent, because of either its signaling or risk-sharing effect (Stiglitz and Uy 1996). Government policies that increase the equity of a firm can have immense effects through the power of leveraging. Beyond that, there was a wide range of instruments for effecting industrial policies; it is the cumulative effect of all of these that matters. The criticism is more properly directed at those who have suggested that Japan's Ministry of International Trade and Industry totally controlled the allocation of resources. This assumption is wrong on two counts. First, firms made most of the decisions about resource allocation—influenced, to be sure, by government policies, but not directly controlled by them. None of the East Asian countries is a command-and-control economy. Second, the view that government makes decisions on its own seems misguided. Consultation between business and government was extensive (and many of the top leaders of business were former government employees).

The charge that industrial policy was distortionary, however, is of more concern. Even if there is a rationale for government intervention, this view alleges that government does not do a good job at picking winners. Instances of mis-

takes by the government are typically cited. In some cases the government discouraged a firm (Honda, for example), when in retrospect it clearly should not have; and in others the government encouraged some industry (such as petrochemicals), when in retrospect it probably should not have.

There are four responses to this criticism. First, good decisionmaking by the government necessarily involves making mistakes: a policy that supported only sure winners would have taken no risks. The relatively few mistakes speak well for the government's ability to pick winners. Second, the government was not heavy-handed. Although it made mistakes of judgment, it did not force its opinions on others when they were willing to risk their own capital. This is one of the strengths of decentralized decisionmaking: it ensures that mistaken views will not dominate.

Third, to a large extent, government policies were not directed at picking winners in the narrow sense of the term. Several governments decided to support export-oriented industries. In a sense, that was choosing a winning development strategy; it did not necessarily entail micromanaging. Even when the government identified an industry for support, the banks seem to have had discretion to select which firms or projects within that industry to support.

Fourth, industrial policies were focused not so much on picking winners as on identifying market failures—instances where investors could not capture large potential spillovers. Concern about such spillovers helps explain the government's encouragement of high technology industries. Training provides another example. Firms would benefit from a trained labor force, but, because workers can leave for a better job once they are trained, firms have inadequate incentives to proceed with training. Yet a skilled work force is essential for economic growth, so government undertook to improve the quality of the labor force by emphasizing education.

Moreover, the criticism of industrial policies as misguided attempts to pick winners ignores the broader range of government actions, such as its role in spearheading the expansion of certain manufacturing sectors. "Picking winners" seems to imply culling from a fixed pool of applicants to find those with the highest long-run social returns. East Asian governments have instead performed an entrepreneurial role. Entrepreneurship requires combining technological and marketing knowledge, a vision of the future, a willingness to take risks, and an ability to raise capital. In early stages of development, these ingredients are typically in short supply. The governments in East Asia stepped in to fill the gap—but in a way that promoted rather than thwarted the development of private entrepreneurship.

Government was also effective in monitoring the recipients of its support and ensuring that they did not siphon off funds for private use. Other government policies, such as those that led to more equity financing, reduced the magnitude of the monitoring problem; that is, they resulted in firms having more appropriate incentives. Still other policies, such as those that enhanced the stability of the banking system, led to more effective monitoring by financial institutions.

Cooperation and Competition

Popular discussions of the success of Japan and several other East Asian countries have stressed the cooperative relations between government and business, between workers and employers, and between small and large businesses. Clearly, the extent of this cooperation (sometimes referred to as "Japan Inc."), has been exaggerated. Yet a variety of institutions and practices facilitate cooperation, and this kind of cooperation appears to have had beneficial effects. Adam Smith's "invisible hand" of perfect competition argues that because each individual, in pursuing his self-interest, is also maximizing the common welfare, cooperation is not necessary. But when market failures occur, it is not necessarily the case that the selfish pursuit of self-interest leads to efficient outcomes.

The governments of East Asia recognized that the business community had superior information about investment decisions, but they also recognized that the overall information base could be improved. The establishment of formal and informal councils gave rival firms and industries a way to exchange information with each other and with the government. (This information exchange process is sometimes described as akin to indicative planning, but the analogy is, at best, an imperfect one.) These exchanges conveyed far more information than the traditional format used to display planned sectoral inputs and outputs. What made them more meaningful than such exchanges in other countries? Why would businesses, or government for that matter, tell the truth?

To a large extent, good behavior is induced through long-term relationships and reputations. In the process of development, social sanctions become less effective in enforcing cooperative behavior, but establishing and maintaining alternative bases for cooperative relations may be difficult. The gains from cooperation are based on the perception that the future returns to cooperation exceed the short-run gains that might accrue from the pursuit of self-interest. But an environment of rapid change may heighten uncertainty about the value of the future relationship and the magnitude of the long-term gains from cooperation. Moreover, future cooperative gains have to be discounted (meaning that, because they may not materialize in the future, they are worth less in today's terms). Typically uncertainty is greater and discount rates higher in developing countries. Further, concerns about the potential bankruptcy of one or the other firm, which could terminate the relationship, heighten the likelihood that cooperative relationships will not materialize. Under these circumstances, future gains from cooperation must be greater to compensate firms for sacrificing the short-run gains from self-interested behavior.

Encouraging Cooperation

The Japanese government used both carrots and sticks to encourage cooperation and the exchange of truthful information. Although cultural characteristics are often credited with facilitating this harmonious result, other countries with

similar cultural backgrounds have not displayed the same sort of cooperative behavior seen in Japan. It seems far more likely that government actions were more important than culture in shaping these behavioral patterns.

Of the institutions and mechanisms that facilitated cooperation, an important role was played by business councils set up to share reliable and timely information. Why did not some businesses try to "free ride," to obtain the information provided by others while providing no real information themselves? The answer is, in part, that they were in a longer-term relationship; a firm that "cheated" would be ostracized from the circle. The fact that the government was included in these circles was important: firms wanted to know what the government was thinking about specific projects or what policy changes were planned. Even if a firm's cooperative instincts went astray, self interest was a strong incentive. Moreover, by paying attention to these councils, the government ensured that the gains from cooperation were even greater.

The government's discretionary powers enabled it to reward cooperation and honesty, and there was at least a fear that the lack of cooperation and the appearance of dishonesty would be punished. Government intervention in markets created rents that the government could then allocate to participants who behaved cooperatively. For example, by restricting the formation of branch banks, a large franchise value was associated with the right to have a branch. Similarly, restricting credit meant that access to credit had value. And the Bank of Japan (the central bank) could, on a discretionary basis, provide banks with additional funds when needed.

The relative stability of the East Asian governments increased the incentives for establishing long-term cooperative relations. At the same time, long-term relations enhanced the effectiveness of incentives (Stiglitz and Weiss 1983). Firms that performed well on one project could expect to be rewarded with another project.

The East Asian governments also tried to create an environment conducive to close cooperation among businesses. In Japan, for instance, the government tried to encourage mergers. To the extent that these programs were successful (and there is considerable controversy about that), they reduced the difficulties of cooperation. The smaller the group, the easier cooperation is to attain. Here the government was walking a fine line; a small group could—and may—have led to collusion by restricting competition. In some circumstances, the government approved the formation of so-called recession cartels. These cartels were an explicit attempt to deal cooperatively (and collusively) with the problems that arise in a recession when there is excess capacity in a capital-intensive industry. Under certain conditions, as demand shifts down, prices drop and firms are unable to recover their capital costs. Recession cartels were a way to restrict competition to enable the industry in question to avoid the low prices that would damage all the firms. Whether the gains were worth the costs of reduced competition, higher prices, and underutilized resources is not clear, however. Because of the strong incentives to cheat on such arrangements, cartels are seldom suc-

cessful without legal sanctions from the government. In some cases the Japanese government paid firms to destroy equipment, and in others, to seal equipment shut. Even these tactics were not always successful; some firms did not completely dismantle their equipment.

Labor markets were similarly designed to encourage cooperative behavior. The Japanese pattern of lifetime employment was important because it meant that employees had long-term relationships with employers, which facilitated cooperative behavior. The rapid increase in wages that came with age and experience provided a strong incentive for workers to stay with their organizations. The average pay of each age cohort increased sharply, but differentiation within the age cohort remained smaller than in, say, the United States. Japan's prevalent compensation scheme, in which a large part of the salary was paid as an annual bonus (based largely on profits of the previous year), also encouraged cooperation because workers had, in effect, an equity stake in the firm. This form of risk sharing may be particularly important in early stages of development when capital markets are underdeveloped. Because wages are based on the group's performance, the individual has an incentive to monitor his peers to make sure that his co-workers are working hard (Arnott and Stiglitz 1991; Stiglitz 1990b). One might even go further. Basing salary on individual performance encourages self-interested, noncooperative behavior. Conversely, paying wages based on group performance signals the importance of cooperative behavior.

Also important in Japan's labor market was the government-established Productivity Council, which dealt with the degree of inequality that could exist within a firm and limited salaries of top managers to no more than ten times the wages of the lowest paid workers. This compressed wage structure enhanced the sense that top management was not taking advantage of workers and led to greater effort and lower labor turnover.

Cooperative behavior between firms and their employees is particularly important in facilitating technological change. Workers are often in the best position to identify improvements in efficiency, although such improvements do not always redound to the benefit of the workers. Because labor saving innovations may result in less demand for labor and higher unemployment, employees are often reluctant to disclose such ideas. If, however, the firm provides a guarantee of lifetime employment, existing employees will see no conflict between their interests and those of the firm. Moreover, when wages are based partially on firm profitability, interests coincide: if the productivity enhancing innovation enhances profits in the long run, employees will share in the gain. Of course, when growth is rapid firms can more easily promise that labor-saving innovations will not result in reduced employment, which makes it more credible that all (existing) employees will benefit from such innovations.

Cooperative behavior between firms and their banks was also evident in the operations of capital markets. In Japan each firm had a long-standing relationship with a single bank, and that bank played a large role in the affairs of the firm. Japanese banks, unlike American banks, are allowed to own shares in the

firms to which they lend, and when their client firms are in trouble, they step in. (The fact that the bank owns shares in the firm means that there is a greater coincidence of interest than there would be if the bank were simply a creditor; see Stiglitz 1985.) This pattern of active involvement between lenders and borrowers is seen in other countries of East Asia and was actively encouraged by governments.

Another important aspect of business-government cooperation in Japan has been the attempt to reduce bankruptcies, which have been markedly less cyclical than those in the United States and other countries. This pattern reflects not only the country's better macroeconomic performance and a legal structure that encourages actions short of bankruptcy but also an active government policy directed at avoiding the economic disruption caused by bankruptcy.

Combining Competition and Cooperation

The East Asian countries succeeded (not always, but with a remarkable frequency) in harnessing the advantages of cooperation while retaining the advantages of competition. Cooperation to increase efficiency can easily be turned into collusion to raise prices and restrict output and entry. Worse still, discretionary powers needed for cooperation can give rise to rent-seeking and corruption. Competition both enhanced efficiency and reduced the scope for abuses of discretionary powers. In fostering a competitive industrial structure, governments looked not so much at the number of firms in an industry, but at the effectiveness of the competition; competition may be more effective with two evenly matched firms than with one firm competing with many small rivals (Nalebuff and Stiglitz 1983).

By the same token, the process of identifying which workers to promote in Japanese firms may be more effective in encouraging competition than is the process in the United States. In Japan, where workers are less mobile, a cohort of workers hired together advances together. They all work hard; they all have to signal that they are committed to the firm; they all remain in the contest. In the United States decisions concerning who is on an upward career ladder often take place earlier. Under that system, incentives may be strong in the early stages of individual careers, but they may be greatly attenuated once these decisions are made. Those who know that they are not going to be "winners" have little reason to work hard.

One method introduced to stimulate competition was the use of contests. Governments rewarded firms that performed well relative to others (such as in exports) by, for example, providing them with access to capital and foreign exchange. In many instances, the value of the prize arose from the government intervention: if the government had not created artificial scarcity of capital or foreign exchange, an increase in availability would have had no incentive effect.

Well-functioning contests are characterized by rules that establish a clear criterion for rewards, such as export performance; specify the nature of the reward (the allocation of credit or foreign exchange); and indicate who will evaluate performance. This system reduced the scope for abuse of bureaucratic discretion at the same time that it provided strong incentives.

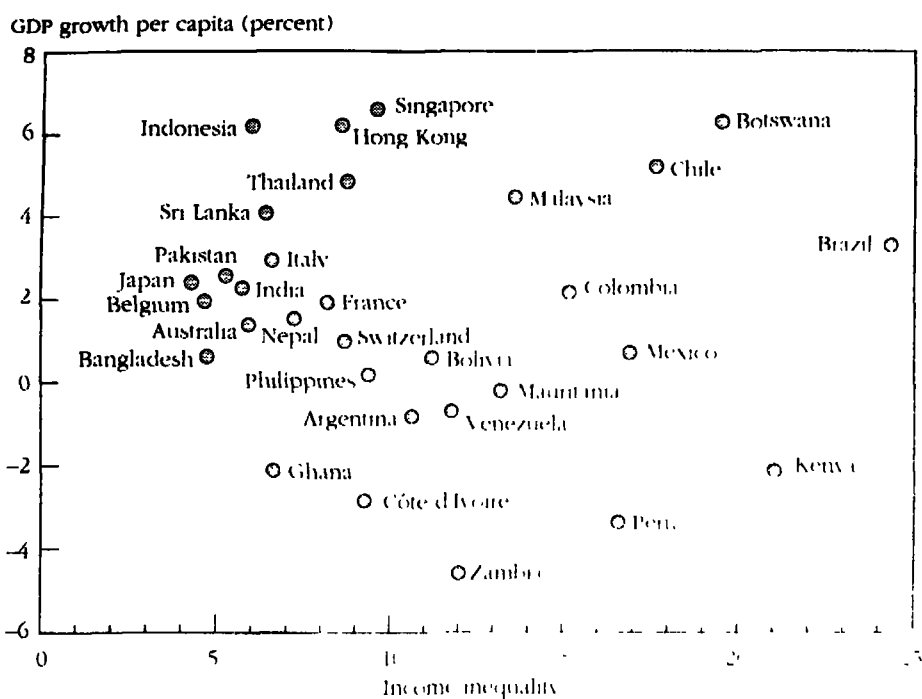
Ironically, licensing requirements put in place to restrict competition may give rise to more competitive behavior. At various times, the Japanese government imposed restrictions on the expansion of capacity in certain industries. It awarded licenses to expand capacity on the basis of firms' previous market shares. Thus performance—particularly growth—in one year may increase profits not only in that year, but also in subsequent years.

Growth with Equality

Although industrial policies attempt to direct resource allocations in ways that maximize growth, income distribution policies seek to promote greater equality. Historically, the development process has been characterized by marked increases in inequality (the Kuznets curve). It was alleged that the massive amounts of capital accumulation required could only be attained through significant inequality; the poor simply could not save enough. Moreover, growth creates winners (the owners of those firms that do well), and losers (workers displaced from lagging industries, in particular agriculture). The economies of East Asia were able to achieve rapid growth without an increase in inequality. Indeed, active policies promoting equality probably enhanced growth (figure 1).

In Korea, Japan, and Taiwan (China), land reforms—at least partially imposed from the outside—were important in the initial stages of development. These had three effects: they increased rural productivity and income and resulted in increased savings; higher incomes provided the domestic demand that was important in these economies before export markets expanded; and the redistribution of income contributed to political stability, an important factor in creating a good environment for domestic and foreign investment.

In later years policies to ensure more equitable distribution of income continued to contribute to economic growth, with positive effects that more than offset the possible negative effects of reduced capital accumulation upon which earlier discussions had focused. These policies continued to contribute to political stability. High and increasing wages reduced inequality, made workers not only more satisfied but also (by standard efficiency wage arguments) more productive, and promoted cooperative relations between workers and firms. Policies that attempted to restrict real estate speculation (by limiting lending for that purpose) can be viewed both as part of industrial policy and as part of income distribution policy. While they directed funds into industry, they limited the increases in the prices of housing relative to what would otherwise have occurred. Such price increases would have led to demands for further wage in-

Figure 1. *Income Inequality and Growth of Gross Domestic Product, 1970–93*

a. Income inequality is measured by the ratio of the income share of the richest 20 percent and the poorest 20 percent of the population.

Source: World Bank data.

creases and would have had particularly adverse effects on the very poor, who often seem unable to obtain adequate housing under such conditions.

Additionally, policies ensuring universal literacy both increased productivity and promoted greater equality. The emphasis on female education led to reduced fertility, thus mitigating the adverse effects of population pressure felt in so many developing countries, and it directly increased the supply of educated labor. Most studies suggest that a worker's wage performance is more directly related to nonschool factors, such as home background, than to education in school. Education of women can be thought of as a roundabout but high-return way of enhancing labor force productivity.

In Thailand a program to provide credit to the rural sector, although largely motivated by concerns about communist insurgency, seemed not only to have promoted equality but also to have yielded reasonably high economic returns. And in Malaysia, policies that would be regarded as affirmative action elsewhere were able not only to draw upon a reservoir of human talent that had not

been well used before, but to weld together a nation that had already demonstrated a potential for ethnic strife.

There are positive relations between growth and equality. High rates of growth provided resources that could be used to promote equality, just as the high degree of equality helped sustain the high rates of growth. Although this may seem to be little more than common sense, until the experience of East Asia, "common sense" suggested quite the contrary: growth produced inequality, and inequality was necessary for growth.

Export-Oriented Growth

Why focus on exports? Should not countries simply produce the goods in which they have a comparative advantage, whether that happens to be products that are exported or substitutes for goods that are currently imported?[5] Success in exporting provided policymakers with an objective way to award credit and foreign exchange.[6] Two questions arise: why are exports a better measure of performance than profits? And second, do markets reward success in an appropriate way without government intervention?

In measuring performance to determine which firms to favor with credit and other scarce resources, governments faced an information problem. All governments face a similar problem, but in the context of development, the information problem is particularly severe for two reasons. First, relatively few firms may be engaged in similar activities, so bases of comparison are limited. Second, a host of problems must be overcome—new supplier relationships, new markets, and so on. In such circumstances, short-run profits may be imperfect indicators of long-term performance. Consider, for instance, the two sources of profits: those derived from exports, and those derived from domestic sales. The latter may reflect either the firm's efficiency or its monopoly position in the economy. The profits that result from imperfect competition in the domestic market accrue at least partially at the expense of consumers and should not be thought of as a social gain. By contrast, a firm that succeeds in the export market is more likely to be economically efficient. It can market a product at a lower price than can foreign rivals, or one better tailored for the world market. Export markets are more likely to be competitive. And even if they are not, it is of no concern: the profits of the firm are then at the expense of foreign consumers. Indeed, from the exporting country's perspective, finding a niche within which some market power can be exercised is to be rewarded, not condemned.

Other advantages are also associated with exports. Firms learn a great deal in international markets, benefiting from spillovers related to both marketing and production know-how. For instance, success in producing intermediate goods requires producing to standards that are typically higher than those that prevail within developing countries. This generates a demand for testing laboratories.

The recognition that standards are important and the knowledge about producing goods of higher quality has implications across a broad range of products. Moreover, the contacts established through exporting may be of value when the firm decides to enter related markets. It will, for instance, know where to turn to acquire advanced technology.

From a social perspective, success in exporting may be a better indicator of whether a firm merits additional funds than success in selling domestically, but banks have typically preferred lending to firms engaged in the domestic market, and for a simple reason. Banks do not care whether a firm makes social returns or private returns, as long as it can repay the loan. Banks are less informed about foreign markets and thus consider it riskier to lend for export projects than for the domestic market.

It has been argued that the preferences East Asian governments gave to exports were intended simply to offset the disadvantages of tariffs and other restrictions on imports. From this perspective, government was not promoting exports but simply "getting the prices right." Upon closer examination (even without a detailed scrutiny of the statistics), this argument appears faulty on two grounds. First, it refers to averages of exports, but what is relevant is the effective subsidy on particular exports. If some exports are encouraged and others (perhaps unintentionally) discouraged, it is apparent that government has intervened in the allocation of resources. Second, the government actually engaged in a wide range of activities beyond direct subsidies to promote exports.

Export Promotion Activities

Four activities were very important in promoting export growth: the provision of infrastructure; preferential access to capital and foreign exchange; the development of export markets; and licensing and other regulations designed to enhance the reputation of the country's exports. As noted previously, the close, long-term relationships between exporters and governments can be credited with making these mechanisms work.

THE PROVISION OF INFRASTRUCTURE. Because poor infrastructure is an important barrier to trade, East Asian governments have invested in infrastructure, including good port facilities and improved transportation systems to reduce the costs of shipping goods abroad. Transportation is not the only aspect of infrastructure that has received government attention. Singapore has been involved in efforts to provide an adequate supply of electricity and an effective telecommunications system, both vital to the country's development as a financial center.

PREFERENTIAL ACCESS TO CAPITAL AND FOREIGN EXCHANGE. Most of the countries of East Asia engaged in some degree of financial restraint, that is, capital markets were controlled to give priority industries preferential access to capital and foreign exchange (Stiglitz and Uy 1996). Although in some cases govern

ments provided subsidies (including lower interest rates) to encourage the expansion of favored industries, most observers believe that the access to credit was far more important.

Critics of this access raise the issue of fungibility: what if the government did provide credit and funds for investment in export-oriented industries? So long as money is fungible, large conglomerates could simply divert to other uses those funds that would have been allocated to exports. Consequently, financing exports may have little incremental effect on exports. From this perspective, the allocation of capital to the export sector has no marginal effect. It has only an inframarginal effect on firms that are successful in exporting. This view, however, does not take account of the process by which funds are allocated. If past export performance is used as one of the criteria for judging the creditworthiness of the borrower, firms have an incentive to increase exports. And firms that were successful exporters had demonstrated some set of abilities. If those abilities were correlated with other abilities that enhanced the likelihood of high marginal returns to investment, then the use of export performance may have been an efficient selection mechanism.

DEVELOPING NEW EXPORT MARKETS. Information problems associated with the development of new export markets go beyond the problems of reputation. One noted earlier was the "public good" nature of information. As in the case of other public goods, a strong case can be made for public provision. And many of the East Asian economies have done just that. For instance, Singapore's Economic Development Board has actively worked on developing foreign markets and takes an active interest in what goods might be produced for export. Business executives are invited to join official trips abroad to persuade them that it is in their interests to enter into meaningful business relationships overseas.

ENHANCING THE REPUTATION OF THE COUNTRY'S EXPORTS. In the 1950s and early 1960s, Japanese products had a reputation for being shoddy. American and European buyers had little information about individual Japanese producers and were likely to make unfavorable inferences concerning any particular product. Because establishing a reputation is expensive for any firm seeking to export (particularly when consumers have strong negative prior beliefs), individual firms had little incentive to improve the quality of their products. The government conducted a concerted effort both to improve the quality of the products and to establish brand reputations for Japanese firms, so that they would have private incentives to maintain their reputation. Here is an example of an interaction between cooperative behavior and individual incentives. A similar process is occurring in Taiwan (China), where the government is effectively providing subsidies for firms to establish brand recognition. In doing so, firms will have a private incentive to maintain high quality, with positive effects on the reputation of Taiwanese products in general.

Conclusion

One of the reasons for attempting to delineate what East Asian governments did that resulted in such high growth rates is that other countries would like to replicate their success. If they did the same thing as the governments of East Asia, would they too grow at such rapid rates? To be sure, many countries did similar things, but often with adverse rather than positive effects. They created development banks, only to find that the development banks diverted scarce savings into projects with low returns and made investments that did more to line the pockets of politicians than to raise the welfare of the country. The East Asian miracle had many dimensions: rents were created, but they were used to encourage growth, not dissipated in rent-seeking. Government and businesses cooperated closely, but they collaborated without collusion. Many aspects of this transformation can be explained, and to the extent that they can be explained, it is possible that what they did can be replicated. A high rate of saving leads to high growth; allocating resources on the basis of contests and other performance-based measures can both provide high-powered incentives and reduce the scope for corruption; egalitarian policies, including active education policies, can contribute to a more stable political and economic environment and lead to faster growth through a more productive labor force. Governments that use markets and help create markets are likely to be more successful in promoting growth than governments that try to replace markets.

What generalizations can be drawn from the findings of this article? To be sure, not all of these generalizations are held with the same degree of confidence. In some cases, there are alternative interpretations of the events and evidence. But a combination of theory and evidence supports these conclusions. Included in the discussion below are several interventions in the financial market, which, although mentioned only briefly, are amplified in the accompanying article by Stiglitz and Uy in this issue. Because governments in different countries pursued somewhat different policies, not all the statements hold with equal validity in all countries; some may not even hold within all sectors of a given country. These conclusions are organized around six themes.

- *Making society function better.* Economic growth required the maintenance of macroeconomic and political stability. Policies that sustained a more equitable distribution of income—and that supported basic education for women as well as men—contributed to economic progress by encouraging political stability and cooperative behavior within the private sector. The result was a better business climate for investment and more effective use of human resources.
- *Adaptability of government policies.* Government policies adapted to changing economic circumstances, rather than remaining fixed. As the East Asian economies grew more complex, government had less need to assume an active role and found it more difficult to act effectively on a broad scale.

- *Government and markets.* Governments played an active role in creating market institutions, such as long-term development banks and capital markets to trade bonds and equities, and in establishing an institutional infrastructure that enabled markets to work more effectively. These institutions and markets helped ensure that the high volume of savings was invested efficiently. Governments also used their control of financial markets to help direct resources in ways that stimulated economic growth. This control was probably more important than direct subsidies or low interest rates. Credit was directed not only toward priority areas, but away from speculative real estate and consumer durables.

  Policies to improve government-business cooperation enabled governments to design programs that served the needs of the business community, created a favorable business climate, and encouraged business to direct its energies in ways that contributed to high social returns. Sharing information enhanced the quality of decisionmaking.

  By using, directing, and supplementing markets rather than replacing them, the private sector remained the center of economic activity in most of the East Asian countries; when the private sector disagreed with the government, it was permitted to go ahead and risk its own capital.
- *Promoting accumulation of physical and human capital.* The introduction of postal savings institutions and provident funds resulted in higher domestic savings. At the same time, measures that established prudential regulations (and in some cases, entry restrictions) enhanced the safety and soundness of financial institutions and promoted financial deepening. A variety of programs increased the returns to private investment and facilitated the development and transfer of technology; these included policies that promoted education and training, provided infrastructure, and, in most countries, established a receptivity to foreign investment.
- *Altering the allocation of resources.* Governments in East Asia used industrial policies to affect the allocation of resources in ways that would stimulate economic growth. They took an entrepreneurial role in identifying industries in which research and development would have high payoffs. Support for industry, such as the establishment of research and science centers and quality control standards, was important both in attracting foreign investment and in encouraging domestic investors. Emphasizing industries with strong backward and forward links and large externalities may have helped long-term growth. In the short term, the lack of profitability does not provide a good measure of the potential long-run contribution to growth, precisely because it is the discrepancy between private and social returns that motivates government intervention.

  Governments actively encouraged firms to export. Exports provided a performance-based criterion for allocating credit, encouraged the adoption of international standards, and accelerated the diffusion of technology. Contests among exporters were used widely as incentive devices. The essential

ingredients of contests are rewards (here the allocation of credit), rules (measures of performance), and referees (who evaluate performance). In a world short of perfect competition, contests can provide strong incentives with limited risks, and, if the rules are well specified, reduce bureaucratic abuses.

- *Government policies supporting investment.* Mild financial repression had a positive effect on economic growth. The effects on national savings and on the efficiency with which scarce capital was allocated were likely positive; positive incentive effects may have been associated with the contest for scarce credit, and the increased equity of firms and banks (because of lower interest rates) enhanced their ability to bear risks. Equally important were other government programs that led to more effective risk-sharing within the economy. Risk-sharing reduced the effective cost of capital, thus stimulating investment. Government intervention in international economic relations (for instance, in bargaining for foreign technology, in impeding certain capital movements, and in insisting on certain transfers of technology as part of foreign investment) may have enhanced the national interest, promoted economic stability, and enhanced savings.

No single policy ensured success, nor did the absence of any single ingredient ensure failure. There was a nexus of policies, varying from country to country, sharing the common themes that we have emphasized: governments intervened actively in the market, but used, complemented, regulated, and indeed created markets, rather than supplanted them. Governments created an environment in which markets could thrive. Governments promoted exports, education, and technology; encouraged cooperation between government and industry and between firms and their workers; and at the same time encouraged competition.

The real miracle of East Asia may be political more than economic: why did governments undertake these policies? Why did politicians or bureaucrats not subvert them for their own self-interest? Even here, the East Asian experience has many lessons, particularly the use of incentives and organizational design within the public sector to enhance efficiency and to reduce the likelihood of corruption. The recognition of institutional and individual fallibility gave rise to a flexibility and responsiveness that, in the end, must lie at the root of sustained success.

Notes

Joseph E. Stiglitz is chairman of President Clinton's Council of Economic Advisers, on leave from Stanford University, where he is professor of economics. This is a shortened version of a paper written as part of the World Bank project on The East Asian Miracle and Public Policy. Financial and technical support of the World Bank is gratefully acknowledged. The author is particularly indebted to Marilou Uy. He has also benefited from discussions with Nancy Birdsall, John Page, Richard Sabot, Howard Pack, Edward Campos, Masahiro Okuno, Masahiko Aoki, Daniel Okimoto, Lawrence Lau, Professor Gato, Professor Baba, and dozens of other government officials, academics, bankers, and

industrialists who gave generously of their time during this research project. Research assistance from Thomas Hellman is also gratefully acknowledged.

1. In the literature on this subject, particular reference should be made to the work of Alam (1989), Aoki (1988), Wade (1990), Amsden (1989), Okimoto (1989), Lau (1990), Agrawal and others (1992), Johnson (1982), Pack and Westphal (1986), Itoh and others (1984), Komiya, Okuna, and Suzumura (1988), and Vogel (1991), as well as to the country studies of the World Bank. The information theoretic foundation of the analyses presented here is set forth in greater detail in Greenwald and Stiglitz (1986, 1988, 1992), Arnott, Greenwald, and Stiglitz (1993), and Stiglitz (1994b). The implications for government policy are discussed in greater length in Stiglitz (1990a, 1991a, 1991b).

2. The contrast between India and Singapore could not bring this point home more clearly. India, with a population 300 times that of Singapore and a gross domestic product ten times as large, has a cumulative foreign investment one-fifteenth that of Singapore's.

3. The discussion of this section focuses on standard market failures associated with externalities, missing markets, and competition. The Greenwald-Stiglitz theorems, which go beyond these standard market failures, establish that whenever information is incomplete, a discrepancy may exist between social and private returns. An important application of this principle arises in the context of capital markets: the ratio of the private return to the supplier of capital to the social return may differ markedly (even in the absence of the traditional market failures). For instance, private lenders may be able to appropriate a larger fraction of the total returns to real estate lending than to other lending. For a fuller discussion of the implications, see Stiglitz and Uy (1996).

4. This should be contrasted with the United States, where, for instance, in recent years top executives often received 100 times the pay of recent hires. Within rapidly growing areas of China, the degree of inequality is even lower, with managers getting paid approximately three times the amount received by workers.

5. Note that several of the countries went through an import substitution phase, during which they were very successful. It is questioned whether this phase was necessary, whether it helped (or hindered) the growth process, or whether it was primarily a consequence of the particular economic doctrines that were fashionable at the time.

6. The arguments here are not those provided by government officials at the time (or even subsequently). These focused on more immediate concerns: for instance, in the postwar era, with an overvalued foreign exchange rate, Japan was short of foreign exchange. To some extent, it saw export activities as offsetting the disadvantages exporters faced as a result of the overvalued exchange rate.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Agrawal, P., S. Gokarn, V. Mishra, K. Parikh, and K. Sen. 1992. "Learning from Tigers and Cubs." Discussion paper. Indira Gandhi Institute of Development Research, Bombay. Processed.

Alam, M. S. 1989. *Governments and Markets in Economic Development Strategies.* New York: Praeger Publishing.

Amsden, Alice H. 1989. *Asia's Next Giant.* New York: Oxford University Press.

Aoki, Masahiko. 1988. *Information, Incentives, and Bargaining in the Japanese Economy.* Cambridge, U.K.: Cambridge University Press.

Arnott, Richard, Bruce Greenwald, and Joseph E. Stiglitz. 1993. "Information and Economic Efficiency." NBER Working Paper 4533. National Bureau of Economic Research. Cambridge, Mass. Processed.

Arnott, Richard, and Joseph E. Stiglitz. 1991. "Moral Hazard and Non-Market Institutions: Dysfunctional Crowding Out or Peer Monitoring." *American Economic Review* 81(March):179–90.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." In National Bureau of Economic Research, *The Rate and Direction of Inventive Activity: Economic and Social Factors.* Princeton, N.J.: Princeton University Press.

Arrow, Kenneth, and Gerard Debreu. 1954. "Existence of Equilibrium for a Competitive Economy." *Econometrica* 22:265–90.

Carroll, Chris, David N. Weil, and Lawrence H. Summers. 1993. "Savings and Growth: A Reinterpretation." Paper presented at the Carnegie-Rochester Public Policy Conference, Bradley Policy Research Center, Rochester, N.Y. April 23–24. Processed.

Dasgupta, Partha, and Joseph E. Stiglitz. 1988. "Learning by Doing, Market Structure and Industrial and Trade Policies." *Oxford Economic Papers* 40:246–68

Greenwald, Bruce, and Joseph E. Stiglitz. 1986. "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics* 101(May):229–64.

———. 1988. "Pareto Inefficiency of Market Economies: Search and Efficiency Wage Models." *American Economic Review* 78(2):351–55

———. 1992. "Information, Finance, and Markets: The Architecture of Allocative Mechanisms." *Journal of Industrial and Corporate Change* 1(1):37–63

Itoh, Motoshige, Kazuharo Kiyono, Masahiro Okuno-Fujiwara, and Kotaro Suzumura. 1984. "Economic Analysis of Industrial Policy." *Kikan Gendai Keizai (Contemporary Economy)*. Tokyo.

Johnson, Chalmers H. 1982. *MITI and the Japanese Miracle.* Palo Alto, Calif.: Stanford University Press.

Kim, Jung, and Lawrence J. Lau. 1993. "The Sources of Economic Growth of the East Asian Industrialized Countries." Paper presented at a Conference on the Economic Development of Republic of China and the Pacific Rim in 1990 and Beyond. Stanford University, Calif. Processed.

Komiya, Ryutaro, Masahiro Okuno, and Kotaro Suzumura, eds. 1988. *Industrial Policy of Japan.* San Diego: Academic Press.

Krugman, Paul. 1994. "The Myth of Asia's Miracle." *Foreign Affairs* 73(November/December):62–78.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45(1):1–28

Lau, Lawrence J. 1990. *Models of Development: A Comparative Study of Economic Growth in South Korea and Taiwan.* San Francisco: Institute for Contemporary Studies Press.

Murphy, K. M., A. Shleifer, and R. W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97(5):1003–26.

Nalebuff, Barry, and Joseph E. Stiglitz. 1983. "Prizes and Incentives: Towards a General Theory of Compensation and Competition." *Bell Journal* 14(1):21–43.

Okimoto, Daniel. 1989. *Between MITI and the Market.* Palo Alto, Calif.: Stanford University Press.

Pack, Howard, and L. E. Westphal. 1986. "Industrial Strategy and Technological Change." *Journal of Development Economics* 22:87–128.

Rosenstein-Rodan, P. N. 1943. "Problems of Industrialization in Eastern and South-Eastern Europe." *Economic Journal* 53(June-September):202–11

Solow, Robert M. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39 (August):312–20

Stiglitz, Joseph E. 1974. "Theories of Discrimination and Economic Policy." In George M. von Furstenberg and others, eds., *Patterns of Racial Discrimination.* Lexington, Mass.: D.C. Heath.

———. 1985. "Credit Markets and Control of Capital." *Journal of Money, Banking and Credit* 17(1):133–52

IN SEARCH OF OWNERS: PRIVATIZATION AND CORPORATE GOVERNANCE IN TRANSITION ECONOMIES

Cheryl W. Gray

Experiments in privatizing enterprises in transition economies abound, from extensive efforts at sales to strategic owners (as in Estonia and Hungary), to programs based primarily on insider buyouts (as in Russia and Slovenia), to innovative mass privatization programs involving the creation of large and powerful new financial intermediaries (as in the Czech and Slovak republics and Poland). Each approach has inherent strengths and risks. But if the objectives are to sever the links between the state and the enterprises, to school the population in market basics, and to foster further ownership change, the initial weight of evidence seems to favor significant reliance on voucher privatization, especially given the difficulty most countries have finding willing cash investors.

Socialism's defining characteristic was state ownership of all productive assets. Moving from state to private ownership and creating the conditions to improve the performance of medium-size and large enterprises are the main tasks of the transition from socialism. But privatization goes beyond a change in ownership. Programs to privatize enterprises in transition economies should be evaluated in terms of three broad dimensions: the corporate governance mechanisms they create, the supporting institutions they foster, and the extent to which they create a self-sustaining economic and political reform process. Although some governments judge revenues to be an important goal of privatization, they are at best a secondary objective.

The initial patterns of ownership resulting from any program of enterprise privatization are unlikely to be optimal. They may be too dispersed, or they may be concentrated in the hands of entities unable or unwilling to use such resources efficiently. The long-run success of any privatization program therefore depends on the capacity for ownership patterns to change and evolve

The World Bank Research Observer, vol. 11, no. 2 (August 1996), pp. 179-97

to more efficient forms. Programs that spur institutional development, particularly the growth of capital and asset markets, will have a distinct advantage in this regard.

Goals of Privatization

Creating "Real" Owners

The primary economic goal of enterprise privatization is to create true representatives of capital—individuals or groups of individuals who clearly reap the gains from improved performance. Transferring property rights to new owners is just the first step, however, and on its own is not sufficient to guarantee changes in the behavior of managers. The new owners must also have the power, incentive, and ability to monitor managers and ensure that they act in the owners' best interests. Typically, in small firms the owners and managers are one and the same; in large firms, however, ownership and management are usually separate, creating the need for monitoring.

Shareholder monitoring is only one of numerous constraints on managerial behavior in market economies, but it is likely to be more important in the early stages of reform in Central and Eastern European economies where markets for products, capital, and managerial talent are underdeveloped and may not exert strong competitive pressures on managers. Shareholder monitoring can be passive or active. Passive shareholders simply sell their shares to discipline managers, while active shareholders exert their views and vote their shares. In countries where stock markets are still in their infancy, passive monitoring is unlikely to be efficient, and thus active shareholder monitoring is likely to be a critical mode of corporate governance in the near term. Furthermore, under socialism production was so inefficient that major improvements in efficiency are likely to depend more on restructuring than on marginal changes in managerial behavior. Alternative patterns of corporate governance should therefore be judged on how they affect not only day-to-day decisionmaking, but also a firm's capacity for radical change and restructuring.

Changes in corporate governance are not the only potential benefits that come with real owners. Privatization can also stimulate an infusion of capital, technology, ideas, and skills, complementing changes in incentives and boosting the productivity of enterprises. Whether these benefits arise depends to a large extent on the technique that is used to privatize and the distribution of ownership that results. Different types of private owners—whether "insiders" or "outsiders," individuals or institutions, residents or foreigners—all bring different mixtures of goals and capabilities to the firms they own. In some cases, the move from public to private may involve intermediate forms of property—neither wholly public nor wholly private—with their own rationality in the particular setting and their own distinct incentive characteristics (Stark 1996).

Developing Supporting Institutions

Although many market economies have pursued privatization since the early 1980s, the formerly socialist economies face an especially formidable task. They must not only change ownership, but also establish the institutions of a private market economy. Socialism either crippled or reoriented these institutions to reflect the goals of central planners. Legal frameworks defining property rights, private contract regimes, fiduciary liability, dispute resolution mechanisms, and rules of entry and exit for private firms atrophied. Courts lost much, if not all, of their independence as well as their role as adjudicators of commercial disputes and enforcers of commercial laws. Banks lost their independent monitoring role over firms and became instead passive funnels for channeling state funds. "Watchdog" institutions that provide critical information for markets to function, such as credit-rating and consumer protection services, accounting and legal professions, and independent journalism, had neither reason nor permission to exist. Finally, socialism inhibited (indeed, often classified as illegal) the development of basic norms and ethics of market conduct and fiduciary responsibility on which so much behavior in advanced market economies implicitly rests. These laws, organizations, professions, and commercial norms must now be rebuilt, sometimes from scratch.

Creating a Sustainable Reform Process

Transforming property rights and building the institutions of a private market economy necessarily take time. These reforms must therefore be politically and economically sustainable and mutually reinforcing. Yet the often-profound tension between preexisting patterns of state enterprise control and reformers' desire to ensure a positive economic outcome can complicate privatization.

On the one hand, experience shows that the design of a privatization program must take into account the interests of and distribution of power among existing stakeholders—that is, anyone who has an interest in the enterprise, whether economic or political, including managers, employees, and government bureaucrats. Incentive and efficiency problems were pervasive in all socialist economies, but the distribution of power among stakeholders varied from country to country. Earlier reforms toward "market socialism" in Poland, the former Yugoslavia, and to some extent Hungary gave rank-and-file employees extensive powers to influence decisionmaking. In contrast, employees had very little power in East Germany or the Czech and Slovak Federal Republic (the former Czechoslovakia), where control remained firmly in the hands of management, government ministries, and the party bureaucracy until the demise of socialism, after which it shifted to the new democratic leadership. The situations in Bulgaria and the former Soviet Union were somewhere between these two extremes; some influence had devolved to workers, but bureaucrats and managers retained strong powers. Thus the Czech and Slovak Federal Republic and East Germany

could design and effectively implement top-down privatization programs; Poland, Russia, and Slovenia had no such option.

On the other hand, accommodating stakeholder interests clearly has its risks. Compromises that are made to co-opt stakeholders or overcome informational or institutional weaknesses may have negative economic or political repercussions down the road and may undermine long-term economic and political stability. That may occur, for example, if newly privatized firms fail to restructure because new owners lack the incentives or skills for effective corporate governance, if the public perceives privatization as corrupt or highly inequitable, or if privatization concentrates economic and political power in the hands of a small domestic or foreign elite rather than an expanding, independent middle class.

Recent events illustrate this conflict between what is "doable" and what is optimal. In Russia, for instance, the preferences initially given to managers to garner political support for the program are proving costly. Not only is there limited evidence that managers are restructuring existing (largely insider owned) firms, but resentment is growing over the concentrated wealth and power that has resulted from privatization. In the Czech Republic, the large state banks were encouraged to set up investment funds as a way to solidify public support for privatization (in part because they, particularly the savings banks, were among the more trusted institutions at the time). Yet this decision may prove counterproductive in the longer run because extensive economic and political power has come to rest in the hands of a few banks and funds, themselves linked to the government through both formal and informal ties.

In adapting privatization strategies, certain early steps appear to increase the sustainability of reform in any setting. Countries with legacies of strong bureaucratic control over domestic industry should move quickly to sever the old links between firms and line ministries—to cut pervasive subsidies, to weaken the ministries' control, and perhaps to abolish branch or sector ministries altogether. Old political links can be cut as well by barring former Communists from government service for a period of time, as was done in the Czech and Slovak Federal Republic. This first step should be accompanied by the quick adoption of a privatization strategy and systematic efforts to prevent the wholesale looting of the newly freed state firms before they are privatized. This is likely to require a combination of "carrots" (linking the future well-being of managers to the quality of the assets they deliver to the eventual private owners) and "sticks" (imposing strong penalties on managers who divert state assets). These actions may be easier to take in the period of "extraordinary politics" immediately following a political break with the past regime (Balcerowicz 1994). After these steps are taken, each stage of the process should ideally create the momentum and incentives for further progress. That occurs, for example, if the new owners lobby politicians to design and implement laws (such as corporate and securities regulations) that further refine and protect their new rights.

Although steady progress is important for momentum, sustainability is not necessarily correlated with speed. Very rapid privatization was pushed

in the former Czech and Slovak Federal Republic and Russia, in large part on the theory that breaking the links with the state was the primary hurdle and that the political window of opportunity had to be seized quickly. Many aspects of these programs are impressive, and they may yet prove to be major success stories, but the initial design decisions taken to ensure speedy implementation have also produced serious problems, particularly in Russia. Their eventual economic and political impacts are far from clear. In contrast, slower privatization programs, such as those in Poland and, to a lesser extent, Hungary, run the risk of barely getting off the ground when political receptivity is greatest and thus of stagnating before major progress can be achieved. They could nevertheless prove successful if steady progress continues. The slowest movers, such as Belarus, Bulgaria, Romania, and Ukraine, run the greatest risk. Because transition governments are weak, the failure to commit to a formal privatization program leaves the door open for managers to strip assets or direct income flows. The economic injustice of such spontaneous privatization may eventually lead to a political backlash that will undermine further reform.

Finally, reforms in property rights must be complemented by supportive reforms in other areas. Fiscal and monetary policies should be refined to foster a stable price system and impose hard budget constraints on firms. One important policy lesson is that any privatization strategy is likely to be faster and work better where government subsidies are limited, inflation is controlled, and markets exert hard budget constraints on firms. If governments continue to soften budget constraints even for private firms, as they do in many transition economies, the purported benefits of privatization (particularly with regard to incentives) may disappear. On the microeconomic side, the strategy calls for reforms in product markets, more competitive labor markets, and institutional reforms to build effective financial markets. All three are needed to complement shareholder efforts to discipline managers.

Methods of Privatization

Privatizing large enterprises has proven more difficult than most observers originally envisioned. Not only are the goals complex and sometimes at odds with each other, but the firms are often ill-suited to the needs of a market economy. Many are overstaffed and inefficient. Reflecting socialism's efforts to make enterprises the providers of social assets as well as income, many are vast conglomerates with housing, medical services, and child care facilities. Because central planners wanted to economize on transaction costs, many are monopolies. Table 1 summarizes the various methods used to privatize medium and large firms and estimates the extent of privatization in seven countries under each method. What lessons have emerged to date?

Table 1. *Privatization Techniques for Medium-Size and Large Enterprises, 1995*
(percent)

Country	*Sale to outside owners*	*Management-employee buyout*	*Equal-access voucher privatization*	*Restitution*	*Other*[a]	*Still in state hands*
Czech Republic						
By number[b]	32	0	**22**[c]	9	28	10
By value[d]	5	0	**50**	2	3	40
Estonia[e]						
By number	**64**	30	0	0	2	4
By value	**60**	12	3	10	0	15
Hungary						
By number	**38**	7	0	0	33	22
By value	**40**	2	0	4	12	42
Lithuania						
By number	<1	5	**70**	0	0	25
By value	<1	5	**60**	0	0	35
Mongolia[f]						
By number	0	0	**70**	0	0	30
By value	0	0	**55**	0	0	45
Poland						
By number	3	**14**	6	0	23	54
Russia[e]						
By number	0	**55**	11	0	0	34

Note: Boldface numbers show the dominant method in each country. Data are as of the end of 199[illegible]
a. Includes transfers to municipalities or social insurance organizations [illegible] through insolvency proceedings.
b. Percentage shares of the number of all formerly state-owned firms, including parts of firms restructured before privatization.
c. Includes assets sold for cash as part of the voucher privatization program [illegible]
d. Percentage shares of the value of all formerly state-owned firms. Data for Poland and Russia are unavailable.
e. Does not include some infrastructure firms. All management buyouts were part of [illegible] tenders. In thirteen cases citizens could exchange vouchers for minority stakes [illegible] investor.
f. Mongolia has used *only* voucher privatization to privatize medium [illegible] enterprise was sold for vouchers, first in a limited closed subscription [illegible] outsiders via the stock exchange (Korsun and Murrell 1994).
Source: Blasi (1994); Korsun and Murrell 1994; [illegible] 1994; [illegible]

Sales to Outside Investors

Before the transition process got under way in earnest, most countries of Central and Eastern Europe that wanted to privatize state enterprises sought to sell them as going concerns. They were following the only known experience at the time, most notably Great Britain and Chile, where privatization through individual sales had been successful. Because capital markets were undeveloped in the transition economies, most countries hoped to sell the bulk of state enterprises directly to large outside investors, generally strategic investors with special

ized knowledge of the industry. Such trade sales were perceived to have three advantages: they would bring in revenue; they would result in real owners who had the knowledge and incentives to govern companies efficiently and the capital to restructure them; and the conditions of the sale could theoretically be manipulated to take special needs into account.

Although these advantages have indeed been evident in some cases, sales to outside investors have proven far more difficult than originally anticipated. Such sales can work when market institutions are in place, but they are problematic when such institutions are in their infancy. East Germany successfully privatized virtually all of its state enterprises through sales to outside investors, but only with massive amounts of political will and technical and financial assistance from West Germany. Among the transition countries, only Estonia and Hungary have managed to privatize the bulk of their state enterprises through direct sales. Poland and Romania pursued sales vigorously in their early efforts at privatization but with limited success.

THE DISADVANTAGES—LESSONS OF EXPERIENCE. What are the disadvantages of this approach? First, the process is hampered by the limited amount of private capital (particularly domestic capital) available and by the poor quality of information about the condition of the enterprises. One option, which was followed widely in Hungary, is to sell to foreign investors who have sufficient capital and are willing to incur risks or to invest in information-gathering that might decrease such risks. This somewhat controversial strategy may nonetheless be necessary if trade sales are to succeed, and even then, of course, there will be many state enterprises that overseas investors have no interest in buying. Foreign interests have tended to concentrate in certain sectors, such as automobiles, food processing, tobacco, and certain consumer products, whose international market structure tends to be dominated by large, oligopolistic firms (Kogut 1996). A second option is to require less capital up front, so that owners can pay in installments, perhaps out of the future earnings of the firm. Variants of this approach have been tried, for example, in Estonia, Hungary, and Poland. Enforcing future payments is often difficult, however, and renationalizing firms if such payments are not made is equally problematic.

A second disadvantage is that both the process and the resulting distribution of ownership rights may be perceived as unfair. Not only are ordinary citizens unable to participate, but individual deals often look arbitrary, if not corrupt. This perception has been most notable in dealings with foreign investors, with whom packages of incentives and legal regulations have often been negotiated case by case.

Third, the approach tends to be costly and slow, due to the sheer magnitude of the job of evaluating and negotiating deals for each company one-by-one and of providing follow-up monitoring to ensure that the buyers fulfill contract provisions. Sales have also been slowed by other uncertainties, such as who is responsible for cleaning up past environmental damage and how to compensate

those who owned the property before the socialist government came to power (Rutledge 1995).

Fourth, the process is complicated and often stymied completely by the difficulty of placing a value on firms to be offered for sale. Accounting standards inherited from socialism were inadequate to determine the historical value of a firm—much less net present value, on which sales price should theoretically be based. Furthermore, price and trade reforms quickly reduce—if not eliminate—the relevance of previous experience. Potential buyers face profound uncertainty about what these companies will look like in the future. What products will they produce, and for what markets? In what quantities and at what costs? What financing will be available? At what interest rate? Given these uncertainties, calculating even the rough value of an investment is virtually impossible in many cases, even with reforms in accounting techniques.

Finally, insiders have used either explicit or implicit power to block sales to outsiders, particularly in countries such as Poland and Russia in which decisionmaking power had been decentralized. Furthermore, the strength of the insiders' incentives to block a sale is likely to be correlated with the potential profitability of the firm itself, and thus it may be harder to sell the better firms—exactly those for which there is likely to be greater demand from outside investors.

These disadvantages have been more debilitating than initially expected. The Treuhandanstalt in East Germany was able to privatize (or liquidate) its 8,500 state enterprises relatively quickly but at an enormous cost in terms of both skilled manpower and explicit or implicit subsidies to buyers (von Thadden 1994). Other countries, which lacked a benefactor of West Germany's economic strength, could move only slowly—or adopt radically different divestiture techniques. In five years (1990–94) Hungary was able to transfer only about one third of its state-owned assets to private hands through formal sales programs (Pistor and Turkewitz 1996); large sales in infrastructure and energy firms late in 1995 pushed this total somewhat higher. With extensive assistance from former Treuhandanstalt officials, Estonia sold most of its 350 enterprises in three years (1992–95). These are the "successful" cases. None of the other countries of Central and Eastern Europe or the former Soviet Union have even come close to these achievements (in part because foreign investors have shown less interest). Overall experience in the region has led observers to conclude that sales, while a useful pillar in the privatization process, cannot be the sole or even the primary instrument in transition economies.

PRIVATIZING THROUGH EQUITY OFFERINGS. Firms can also be sold to the public by floating shares on public stock exchanges, but this approach is limited by the capacity of the infant stock exchanges in transition economies. Furthermore, it tends to work only for the very best firms with good financial prospects and strong reputations. It is not an avenue for restructuring, not only because poorly performing firms are unlikely to be listed successfully, but also because the dis

persed ownership structure that results is unlikely to create incentives for owners to make the necessary changes. Poland has perhaps had the most success with this approach but still has privatized only about two dozen firms in this manner. Initial public offerings are clearly not the answer to the need for rapid and large-scale privatization, but on the margin they can help develop capital markets.

Management-Employee Buyouts

A second important method of privatization is the sale or giveaway of all or part of the company to its managers and employees. Croatia, Georgia, Poland, Romania, Russia, and Slovenia have all relied primarily on management-employee buyouts, sometimes using government-issued vouchers to provide the liquidity for insiders (and a few outsiders) to purchase shares. To speed up privatization, Hungary implemented a similar small program in 1993 to supplement its emphasis on trade sales. Although the governments of Lithuania and Mongolia did not originally set out in this direction, many firms divested through voucher privatization programs became, in effect, management-employee buyouts (Korsun and Murrell 1994).

An important advantage of this approach is its feasibility and political popularity. In countries in which insiders already had extensive power, those insiders have generally been able to retain their influence during the transition period and effectively maintain a veto power over privatization decisions. In some countries this veto power is explicit; in Poland, for example, employee approval is required for a privatization plan to go forward. In most countries, however, such veto power is implicit. To garner political support, governments have often offered generous benefits to insiders.

A second potential advantage is the one stressed by most proponents of employee share owning plans in advanced market economies. Insider ownership can be both more equitable and, under certain conditions, more efficient (Hansmann 1990; Earle and Estrin 1996; Shleifer and Vasiliev 1996). It can be more equitable because it rewards those who do the work—ironically, the argument at the very heart of socialism. It can be more efficient because it aligns the incentives of owners and workers. Managers and employees may be willing to work harder, monitor each other more carefully, and push for greater productivity if they reap the residual gains.

THE DISADVANTAGES IN TRANSITION ENVIRONMENTS. These potential advantages are counterbalanced by serious disadvantages that are particularly acute in transition settings. First, the process may be inequitable, as employees, particularly those in good firms, reap most of the benefits. One study, for example, calculated that the 19 percent of Russians employed in privatized firms obtained 56 percent of equity sold through June 1994, while the 81 percent of Russians who had only vouchers ended up with only 15 percent of the equity (Blasi 1996).

Second, giving preferences to insiders inhibits—and may even eliminate—competition in the privatization process. Insiders are generally unable to bring new skills and new capital to the company, and socialist managers may have few of the skills needed in a market economy. Because potential outside owners who may be more qualified are not given the chance to participate, the resulting ownership pattern is likely to be suboptimal for the economy as a whole, at least initially. Research in Central Europe confirms that firms privatized to insiders are less likely to restructure and invest than firms sold to outsiders (Earle and others 1994; Barberis and others 1995). Furthermore, changes in ownership patterns may be blocked if managers try to stop employees from selling their shares. Even if employees are free to sell, there may be few buyers. If corporate law and disclosure rules are underdeveloped and thus provide little protection for outside shareholders, as is true in virtually all transition countries (and in some advanced market economies), outsiders may be unwilling to invest in firms with significant insider ownership.

Russia's voucher privatization program, completed in mid-1994, resulted in insider ownership of about two-thirds of the shares of privatized companies. Some managers have tried, albeit illegally, to prohibit workers from selling their shares to outsiders. Others have used less transparent means to block participation by employees or outsiders. For example, managers may attempt to change the form of the company from joint stock to limited liability, because the latter restrains sales of shares to outsiders. Alternatively, they may try to convince employees to put their shares in a trust and assign their voting rights to the managers of the firm. Even when the rights of workers to vote their shares are not restricted, managers may—and do—convince workers that incumbent management is on their side but that outsiders will fire them if they are allowed in. Finally, managers may try to get around employee ownership altogether by setting up new firms and using their inside information and power to transfer valuable assets to these firms. If such a pattern is repeated on a wide scale, this form of ownership may inhibit rather than reinforce the development of a private market economy. It may also backfire politically, if the fruits of privatization become more and more concentrated in the hands of the few, unleashing growing resentment among those ostensibly included at the beginning but ultimately cheated of their expected gains.

ADDRESSING THE PROBLEMS WITH BUYOUTS. How can the advantages of management employee buyouts be enhanced while their disadvantages are mitigated? First, governments need to cut subsidies, liberalize prices, and institute trade reforms to force insider-owned firms to abide by the rules of the marketplace. These steps will help to ensure efficiency regardless of ownership.

Second, government should encourage the entry of new firms to increase competition. Management-employee buyouts may work well for smaller manufacturing and service firms in sectors that are attracting new domestic entrepreneurs; product market competition will keep the insider owned firms on their

toes. Foreign competition could potentially do the same for larger firms, but in such cases managers are more likely to turn to the state to block such competition or to obtain support of one kind or another.

Third, governments should adopt policies that encourage share trading and thereby develop markets for corporate control. In advanced market economies, insider (particularly worker) ownership has an inherent tendency to "degenerate" into investor ownership over time (Earle and Estrin 1996). Whether the same tendency exists in transition environments has yet to be seen. Ownership change requires both a supply of and a demand for shares. To create a supply, shares must be immediately tradable without limitation. To create a demand, outside investors must have not only sufficient capital, but also basic information and protection against fraud and abuse by insiders. These in turn require well functioning corporate laws, securities regulations, and accounting systems. Such shareholder protections do not arise in a vacuum but go hand-in-hand with other economic reforms (Gray and Hendley 1995).

In sum, management employee buyouts excel in their capacity to adapt to the implicit or explicit demands of existing stakeholders, but are less effective in creating corporate governance mechanisms or in attracting new capital and skills. For firms that cannot survive without restructuring, the conflicts of interest that confront insiders may prove particularly unwieldy. In such cases, insiders may look to the state for help and, given political pressures, the state may be willing to listen. This route may thus work better for viable firms that can generate internal funds for investment and may be suitable for small firms without political clout. Indeed, in the case of the latter, employees may be more willing to take painful wage cuts to preserve the company (Earle and Estrin 1996). In the case of large distressed firms with major capital needs, however, this strategy is unlikely to generate the resources, incentives, and capabilities necessary to undertake large scale change.

These pros and cons are particularly apt for firms in which insiders have majority interest. There are, in contrast, strong advantages and relatively few disadvantages to giving insiders minority ownership rights. One clear advantage is that privatization is more likely to be perceived as fair if workers share in any upside gains. Another is that employee-owners can monitor managers or outside owners (such as investment funds or foreign owners with significant minority stakes) who may otherwise have an incentive to loot the firms or stifle competition with other firms under their ownership. These advantages are important, considering the political fragility and the general weakness of watchdog institutions in virtually all transition environments.

Equal-Access Voucher Privatization

A third approach used widely in transition countries is voucher, or "mass," privatization. The government gives away or sells low priced vouchers that can then be used to purchase shares in companies, thereby eliminating the problem

at the core of the sales approach—the shortage of domestic capital. So-called equal-access voucher programs embrace outsiders as well as insiders. This form of privatization has been—or soon will be—implemented in Armenia, Bulgaria, the Czech Republic, Kazakhstan, Lithuania, Moldova, Poland, Romania, Slovakia, and Ukraine.

Well-designed voucher privatization can overcome many of the problems encountered with the various sales techniques, notably the perceived unfairness, the shortage of domestic capital, and the difficulty of placing a value on assets. Because the voucher approach can proceed rapidly, it can simultaneously stimulate the development of market institutions, create new owners, and reorient the interests of existing ones toward further reform. Furthermore, it can speedily cut the links between enterprises and the state that both inhibit restructuring and put fiscal pressures on the state.

The main concern when this method of privatization was first proposed, apart from the question of revenue, was its questionable capacity to develop real owners with proper incentives for effective corporate governance and with access to new capital and skills for restructuring. The concern over corporate governance arose in part from the very notion of vouchers, that is, that one did not value what one did not pay for. More fundamental, however, was the fear that the resulting distribution of ownership would be inefficient and would interfere with the development of strong ownership interests. Experience has shown, however, that a wide variety of ownership patterns can result from voucher privatization. Perhaps more important, such initiatives, if well designed, can stimulate the development of capital markets and stock market trading, thus fostering further ownership change and speeding up the development of corporate control. It can, in effect, privatize the privatization process.

COMBINING SALES AND VOUCHERS. All transition economies have chosen to follow several privatization routes simultaneously, albeit with different emphases. The earliest, biggest, and most successful equal access voucher program to date has been that of the Czech Republic, which has transferred the bulk (in value) of its state enterprises through this route. Mongolia used vouchers to privatize 70 percent of its large enterprises. Romania's 1991 privatization program was much smaller; only 30 percent of the shares of eligible firms were involved. The intention, unrealized and replaced in 1995 by a second and larger mass privatization attempt, was to transfer the remaining shares to strategic owners who could effectively govern and restructure the enterprise. Poland's recent mass privatization plan is smaller still, covering only about 500 companies, or fewer than one-tenth of state owned enterprises.

Larger programs have certain advantages, in that they can include both more firms and a greater diversity of firms. To ensure value to participants and thus gain more political support, while at the same time divesting firms that might not attract cash offers, some of the most profitable firms should be included in the program, along with some of the more marginal ones. Perhaps more impor

tant, larger programs can achieve a greater degree of privatization in a shorter period of time.

DECIDING WHETHER AND HOW A FIRM WILL PARTICIPATE. When a voucher privatization program is announced, who decides whether and in what form a particular firm will participate? As with the size of the program more generally, this decision evolves in large part from the balance of political interests and powers in the particular country. The government of the former Czech and Slovak Federal Republic chose which firms would participate, and the Czech government continued to apply this principle after the country split up. In each case, the central government decided on the mix between voucher auctions and other forms of transfer (primarily sales to strategic investors and restitution to former owners) but based that decision on bids submitted by competing bidders and prepared by them with little government involvement. Thus, the process of designing privatization programs for individual firms was decentralized in a competitive framework, but the final decision process was controlled at the top. This approach, attractive both economically and practically, appears to have worked well in a politically centralized environment where there were no strong inside stakeholders.

Poland and Romania (in its 1991 program) both attempted to follow a more centralized approach by giving the government broad powers to decide which firms would participate and how they would participate. Although this strategy was feasible in Romania, because of the country's strong tradition of centralized power, it was contrary to Poland's diffuse power structure. Indeed, managers and employees of Polish firms have maintained effective veto power over the choice of privatization method.

POOLING OWNERSHIP INTERESTS. If the ownership of shares in a particular enterprise were as widely disbursed as the ownership of vouchers, there would be little corporate governance and probably little progress in reforming enterprises. For this reason many mass privatization programs have encouraged the creation of intermediary institutions to pool ownership interests in particular enterprises. The former Czech and Slovak Federal Republic allowed free entry of private mutual funds, and more than 420 participated in the first wave of privatization. These funds competed with each other to acquire vouchers from the public in exchange for shares in the fund. The funds then invested the acquired vouchers in shares of firms being privatized at auctions. This approach, based on free entry and competition, had the advantage of reducing the state's direct control over the process.

In contrast, the Romanian (1991) and Polish plans called for the government to establish a set number of investment intermediaries, staffed by managers chosen by the supervisory boards appointed by the government. The shares of the intermediaries were then distributed to citizens. No auctions were held. The governments hoped that the intermediaries would actively restructure the firms in their

portfolios and then sell their interests to other investors. That objective has merit, but the danger is that the intermediaries may not be subject to direct market pressures and could end up essentially as government-protected state holding companies. In 1991 Romania's state ownership fund was allocated 70 percent of the shares of each commercial company to be privatized and was directed to divest 10 percent of its holdings a year. After four years it had divested almost nothing and had barely begun the necessary restructuring. Poland's approach did not get off the ground until 1995 (after a delay of three years), and it is too early to judge results.

Although the free entry and competition among funds in the Czech Republic may be arguably preferable to the more bureaucratic approach used in Romania and Poland, creating truly private funds with market-based incentives is proving extremely difficult in any transition setting. The perennial question "who monitors the monitors?" looms over every experiment. This challenge is difficult enough in advanced market economies. It is even more problematic in transition environments, where norms of disclosure and fiduciary responsibility are weak, and watchdog institutions and oversight mechanisms are in their infancy. Breaking the links with the state, although desirable to stimulate entrepreneurship and risk taking, also may mean weakening the capacity to monitor the monitors.

In the Czech Republic, as noted, most of the largest funds were founded by, and are still connected with, the large banks through asset management contracts. These banks in turn continue to be closely connected with the government, both through the sizable stake owned by the state-run National Property Fund and through the government's regulatory powers. Although some nonbank funds quickly established their independence and their potential influence over managers, the bank-affiliated funds appear to be less independent and entrepreneurial. The latter may also face a conflict of interest because the bank lend to the same firms that their funds own. On the other hand, these large funds may be more secure investments than the more entrepreneurial funds which might have an incentive to loot an enterprise or take other actions at th expense of fund or enterprise shareholders. To the extent that several funds ow shares in a firm, they have an incentive to monitor each other (barring collusio among the funds themselves). Although limits on ownership by individual fun tend to discourage active involvement by these owners, the potential for cros monitoring is one advantage of such limits.

The Russian privatization program favored insiders but also allowed the fre entry of private investment funds. Although some 600 funds were formed, th were kept much smaller than the Czech funds, and thus they have far less pow and influence. In the Russian environment, with no legal safeguards, less macr economic discipline, and strong insider control, the goals of the funds are f from clear. Their small size may lead to complex coalitions among or betwe them and other actors in the economy. Some funds appear to have been estal lished primarily for trading vouchers, while others are allied with the manage of individual firms, and still others seem to be involved in seeking subsidies fro

government. Only a minority appear interested in owning and improving the performance of enterprises in the economy (Frydman, Pistor, and Rapaczynski 1996).

Intermediary institutions bring several advantages to voucher privatization programs. At a minimum they aggregate the power of individual vouchers and thereby exercise some monitoring functions associated with ownership. In addition, the free entry of independent intermediaries increases private participation and stimulates competition in the market for corporate control. Finally, observers hope that the funds will become the cornerstones (together with banks or even in place of them) for developing the financial infrastructure that is essential in market economies. But achieving these goals is not easy or automatic. Governments need to consider how they might regulate funds to prevent self-dealing and encourage responsible fiduciary behavior. The involvement of foreign financial experts as fund managers and advisors—one advantage of Poland's approach—can help to strengthen the expertise and norms of conduct within funds.

HOW DO CITIZENS USE VOUCHERS? In the Czech and Slovak Federal Republic, vouchers could be invested either in specific companies or in investment funds. In the Romanian (1991) and Polish programs, in contrast, investing directly in firms was not an option. (In Mongolia and in Romania's newer program, investing in funds is not an option!) In Estonia citizens could use their vouchers to acquire shares in firms or to purchase their homes or land (although relatively few shares were in the end offered for vouchers).

There seem to be no obvious costs—and significant benefits—to allowing wide latitude to investors. Options create competition that can spur funds to greater effectiveness, and they force citizens to become actively involved in voucher investments. In addition, options allow investors to tailor their choices to their own personal risk preferences. Although some people prefer direct investments, funds have proven to be more popular investment vehicles than first expected. When the Czech and Slovak Federal Republic program was set up, most vouchers were expected to be invested directly in firms, but 72 percent of vouchers were ultimately invested in funds.

Furthermore, citizens need not be limited to investing their vouchers. Trading them is also a viable option, and such trading may encourage the emergence of strong, interested owners. If trading is not permitted, immediate rights to trade the shares that are acquired with vouchers is a close substitute. Most of the voucher schemes to date have given some latitude to citizens to sell their interests, whether vouchers or acquired shares. Russia allowed trading in vouchers from the beginning. The former Czech and Slovak Federal Republic forbade secondary trading by citizens in vouchers (although the prohibition was not strictly enforced) but encouraged trading in acquired shares. Such trading has developed rapidly through the Prague and other stock exchanges in the country and through off-exchange transactions.

A somewhat surprising development has been the concentrated ownership and cross-ownership that has emerged from voucher privatization in the Czech Republic. Not only is ownership concentrated in a few funds, but individual funds often own shares of directly competing firms. Furthermore, the funds, together with affiliated banks, are locked in an intricate web of cross-ownership (or sometimes self-ownership) as a result of the privatization of the banks through vouchers (Coffee 1996). Thus banks are insulated from competitive pressures, and the government continues to influence the economy through its 40 percent (or greater) residual holdings of shares in privatized banks.

Organizing auctions. Most voucher schemes rely on auctions to allocate enterprise shares. Voucher auctions can be organized either simultaneously or sequentially (Boycko, Shleifer, and Vishny 1994). The Czech and Slovak Federal Republic (in the first wave) and the Czech Republic (in the second wave) followed a simultaneous approach. The Bulgarian scheme proposed for 1996 is also simultaneous. Other countries, including Georgia and Russia, have generally followed a sequential approach. From an economic perspective the Czech model is more efficient, because all options are known to all bidders at the time of the auction, and the value of a voucher (in terms of purchasing power) does not vary over time, as it does in the sequential model. The simultaneous model, however, is also more complex and costly and may be infeasible in a large country.

Shares in an auction can also be allocated two ways. One is simply to divide the shares on a pro rata basis among bidders, based on the number of vouchers issued. The second is to match the bids against some independent measure of value and distribute the shares only when bids and offers match. The Czech approach, which was a modified version of the latter, required several rounds of bidding to equate demand and supply. The result was arguably fairer but perhaps feasible only because of the relatively small size of the country, the relatively strong central control of the government, and the relatively sophisticated level of understanding in the government and the citizenry. The sale was also facilitated by the country's more stable macroeconomic situation, which meant that inflation was moderate and thus the valuations of firms were more meaningful.

Residual state ownership. Finally, voucher privatization schemes vary in the degree of residual ownership maintained by the state. Romania, for example, privatized only a 30 percent share in each enterprise (in 1991); indeed, some observers question whether this was really privatization at all. The Czech Republic and Poland left significant minority stakes in the hands of government property funds, with a view to using these stakes later to attract strategic investors (or otherwise influence events). The Polish government also had the initial power to appoint the management and supervisory boards of the funds.

If the state is to maintain a stake in firms after privatization, its share should be small and temporary, and its stance relatively passive, although it should con

tinue to monitor the firm to prevent fraud and asset-stripping. Extensive residual state ownership and control can lead to conflicts of interest that diminish—or even nullify—the positive effects of privatization (Pistor and Turkewitz 1996).

Conclusions

Experiments in privatization abound, from extensive efforts to sell to strategic owners to programs based primarily on insider buyouts to innovative programs of mass privatization. These efforts are often complemented by extensive restitution to the former owners of the nationalized property and by smaller programs of bank-led debt equity conversions or public offerings of shares on newly emerging stock markets. Each of these approaches has inherent advantages and risks, and in essence the jury is still out as to which will prove best in the longer run. At present, however, if the objectives are to sever the links between the state and the enterprise, to school the population in market basics, and to foster a perception of fairness, the weight of initial evidence appears to favor equal access voucher privatization, particularly given the difficulty most countries face in finding willing cash investors. Competing and well-monitored intermediary funds are an essential component of this approach.

Experience shows that formal privatization programs are only part of the picture and often only a small part, although they have received most of the attention. Firms are breaking apart and consolidating again from state to private ownership or from one private firm to another (Bogetic and Hillman 1995; Stark 1996). As one Hungarian observer noted, this is the period of "primitive capital accumulation" in the post socialist world. Although formal programs may lay important ground rules, the tremendous economic, legal, political, and even moral uncertainty profoundly affect—and may even overwhelm—most formal efforts at privatization, and it is beyond our ability or insight to know what the final results will be. Both the economic outcomes of these various paths and the efforts to assess them are just beginning to yield insights, and it will be years, if not generations, before a definitive story can be told.

Notes

Cheryl W. Gray is principal economist in the Policy Research Department of the World Bank. The author would like to thank Roman Frydman, Alan Gelb, John Nellis, Lucia Swiatkowski Cannon, and Martha De Melo for comments on earlier drafts.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Balcerowicz, Leszek. 1994. "Poland." In John Williamson, ed., *The Political Economy of Policy Reform.* Washington, D.C.: Institute of International Economics.

Barberis, Nicholas, Maxim Boycko, Andrei Shleifer, and Ira Lieberman. 1995. "How Does Privatization Work? Evidence from the Russian Shops." NBER Working Paper 5136. National Bureau for Economic Research, Cambridge, Mass. May.

Blasi, Joseph. 1994. "Russian Privatization: Ownership, Governance, and Restructuring." In Ira Lieberman and John Nellis, eds., *Russia: Creating Private Enterprises and Efficient Markets.* Washington, D.C.: World Bank.

———. 1996. "Russian Enterprises after Privatization." Paper presented at the Allied Social Sciences Association meeting in San Francisco, January. Processed.

Bogetic, Zeljko, and Arye L. Hillman, eds. 1995. *Financing Government in the Transition—Bulgaria: The Political Economy of Tax Policies, Tax Bases, and Tax Evasion.* World Bank Regional and Sectoral Study. Washington, D.C.

Boycko, Maxim, Andrei Shleifer, and Robert W. Vishny. 1994. "Voucher Privatization." *Journal of Financial Economics* 35(April):249–66.

Coffee, John. 1996. "Investment Privatization Funds: The Czech Experience." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

Earle, John, and Saul Estrin. 1996. "Worker Ownership in Transition." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

Earle, John, Roman Frydman, Andrzej Rapaczynski, and Joel Turkewitz. 1994. *Small Privatization.* London: Central European University Press.

Frydman, Roman, Katharina Pistor, and Andrzej Rapaczynski. 1996. "Investing in Insider-Dominated Firms: A Study of Russian Voucher Privatization Funds." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

Gray, Cheryl, and Kathryn Hendley. 1995. "Developing Commercial Law in Transition Economies: Examples from Hungary and Russia." World Bank, Policy Research Department, Washington, D.C. Processed.

Hansmann, Henry. 1990. "When Does Worker Ownership Work? ESOPs, Law Firms, Codetermination, and Economic Democracy." *Yale Law Journal* 99(8):[illegible].

Kogut, Bruce. 1996. "Direct Investment, Experimentation, and Corporate Governance in Transition Economies." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

Korsun, Georges, and Peter Murrell. 1994. "Ownership and Governance on the Morning After: The Initial Results of Privatization in Mongolia." Working Paper 9[illegible]. Department of Economics, Center for Institutional Reform and the Informal Sector, University of Maryland, College Park. Processed.

Kotrba, Josef, and Jan Svejnar. 1994. "Rapid and Multifaceted Privatization: Experience of the Czech and Slovak Republics." *MOCT-MOST: Economic Policy in Transitional Economies* (Netherlands) 4(2):147–85.

Pistor, Katharina, and Joel Turkewitz. 1996. "Coping with Hydra—State Ownership after Privatization: A Comparative Study of Hungary, Russia, and the Czech Republic." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

Rutledge, Susan. 1995. "Selling State Companies to Strategic Investors: Trade Sale Privatizations in Poland, Hungary, the Czech Republic, and the Slovak Republic." CFS Discussion Paper Series 106. World Bank, Resource Mobilization and Cofinancing Information Services Center, Washington, D.C. Processed.

Shleifer, Andrei, and Dmitry Vasiliev. 1996. "Management Ownership and the Russian Privatization." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

Stark, David. 1996. "Recombinant Property in East European Capitalism." In Roman Frydman, Cheryl Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia.* London: Central European University Press.

von Thadden, Ernst-Ludwig. 1994. "Centralized Decentralization: Corporate Governance in East German Economic Transition." Institut für Volkswirtschaft, Universität Basel, Switzerland.

HOW WELL CAN METHOD SUBSTITUTE FOR DATA? FIVE EXPERIMENTS IN POVERTY ANALYSIS

Martin Ravallion

No one doubts that good data are essential to sound policymaking. Alas, data are invariably faulty. Methodological solutions to data inadequacies have often been proposed and implemented, but they have been tested only rarely. Yet the methods that are used may well determine the direction of policy. For example, the particular survey method used—and the way nonsurvey data are interpreted—may be critical in assessing whether a country's strategy for reducing poverty is working. This article shows how counterfactual experiments can help test the reliability of various methods of dealing with common data problems. Well-designed methods—and they need not be very complicated—can help get around the problem, although it appears that substituting method for data is a long way from being perfect.

Objective data obtained from representative surveys of living conditions are widely used to stimulate public awareness of poverty and motivate government actions to benefit the poor. Yet analysts and policymakers routinely find that these data are deficient in one or more important respects and must find credible methods for dealing with those deficiencies. Various solutions have been proposed that rely on certain regularities in living conditions and use a relatively small number of more easily measured variables—such as membership in certain predefined socioeconomic groups—to infer the missing data.

Hidden differences in living standards may well confound such efforts. Even though the partitions commonly used in assessing poverty—such as land ownership or region of residence—reveal large disparities, these disparities may be weak indicators for targeting the poor. In fact, some recent research indicates that variations *between* socioeconomic groups are often dwarfed by differences *within* such groups (see, for example, Datt and Ravallion 1993; Ravallion and

The World Bank Research Observer, vol. 11, no. 2 (August 1996), pp. 199–221

Sen 1994; and Cowell and Jenkins 1994). So targeting policies that use standard "poverty proxies" may well result in considerable leakage to the nonpoor and highly imperfect coverage of the poor.

Surprisingly little effort has gone into assessing the performance of routine tools of poverty analysis. Yet the methods that are chosen to deal with inadequacies in the data may well influence the conclusions analysts draw, including the implications for policy. For example, in research on Indonesia, it has been found that the method used to establish the poverty lines by region can dramatically alter the structure of the resulting regional poverty profile and hence the regional priorities for alleviating poverty; indeed, regional rankings produced by two common methods were virtually uncorrelated (Ravallion and Bidani 1994). Here the method chosen to make up for missing data would markedly influence the precise policy prescription.

This article evaluates some of these techniques by applying them in situations where the data are available and asking how well they work. Fortunately, not all countries are missing the same data, so it is possible to infer how a particular remedial method would work in a "data-poor" setting by comparing how well that same method performs in a "data-rich" setting. Knowing about the strengths and weaknesses of the many methods that have been used to generate information in data-poor settings can also help researchers make choices in collecting new data. Nationally representative socioeconomic surveys can be expensive; good substitutes, if they can be found, could produce savings.

Evaluating Methodology

There are three generic problems with existing poverty data. First, the specific surveys used are sometimes flawed; there may be problems with the questions asked or the sample selected. Second, it is often the case that too few surveys were conducted or they are incomparable over time. Third, researchers often lack complementary data that are needed to interpret the results; for example, making consistent welfare comparisons from survey data can be difficult if data on prices are not also available.

Ingenious—and sometimes simple—methods of dealing with these data problems can often be devised, although how well they work is unknown. One way to test how well the analyst can circumvent these data problems is to devise a counterfactual experiment. The objective is to try the method in a setting where it is not in fact needed—because the required data are already available—and see how well it performs. The following section outlines five such experiments that test how well methodology can substitute for data.

Several observations about these experiments are in order. First, their aim is not to develop "quick and dirty" methods; in some cases the econometric analysis that is used is quite sophisticated. Second, these experiments share the generic problems of any empirical research. That is, the results are to some exten

data-specific, and their robustness in other settings is not known. And measurement errors may be a problem, as always. The following experiments try to assess how well the selected methods perform against data that meet the prevailing standards of best practice but are unlikely to be perfect.

Third, even without measurement errors, reasonable people may disagree about what should be measured. Here the measure of poverty is defined as a lack of command over market goods. Although this measurement is undeniably important, it is clearly not the only dimension of well-being; command over nonmarket goods, such as some publicly provided services, may be an important omission in conventional measures (for further discussion, see Ravallion 1993). Nor do the experiments reported here constitute a comprehensive list of data problems. For example, Chaudhuri and Ravallion (1994) show the difficulties of inferring standards of living through the use of single cross sectional surveys; and Haddad and Kanbur (1990) study the inequality within households that is often hidden in conventional surveys.

Experiment 1: A Poverty Profile with Missing Prices

Policy discussions about poverty are often informed by a poverty profile that shows how poverty measures vary by location, sector of employment, or some other household characteristic. Unfortunately, deficient data and methods often misinform such discussions. One of the most common—and potentially serious—problems is how to make consistent comparisons of what a given unit of money is worth to poor people in one place versus another.

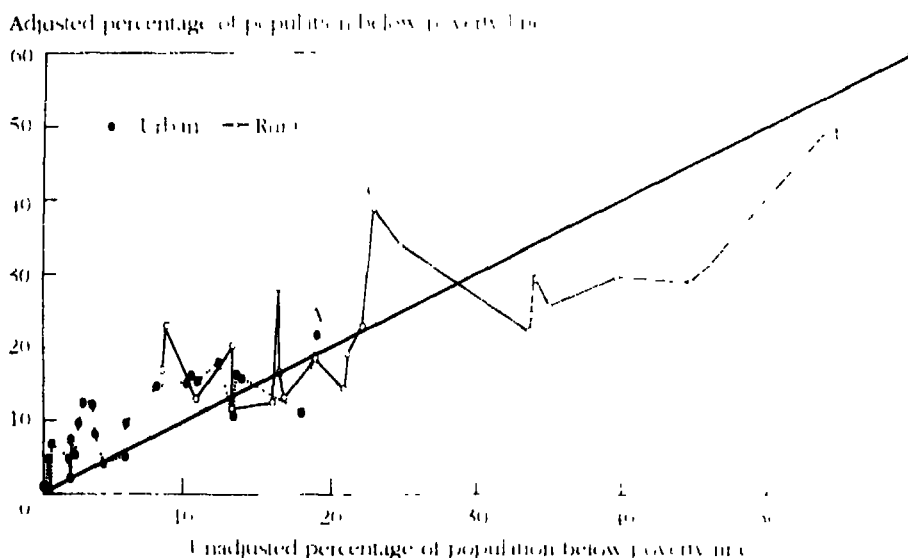
Ideally, studies that compare incomes or expenditures over time or in different regions should adjust for differences in prices. Almost all statistical agencies now monitor prices over time in constructing the Consumer Price Index, but monitoring differences in prices by localities at a given time is far less common. This data gap can be a serious concern in developing countries where the prices of the same goods and services differ widely within the country. Without local price data, one might be tempted to ignore these differences in calculating whether people are poor. How would this omission affect the results? Is an alternative method available that would provide the missing information?

Indonesia is an interesting case study for examining these questions, given its geography and likely locational price differentials. For this experiment, researchers set out to measure the extent of poverty by region and sector of employment, with and without adjusting for differences in the cost of basic needs. Indonesia's Central Bureau of Statistics provided the raw survey and price data. Using the method described in Ravallion and Bidani (1994), the nominal poverty line was adjusted for regional differences in the cost of a bundle of basic foods considered sufficient to meet minimum food-energy (caloric) requirements for good health and normal activities. An allowance for differences in nonfood spending was added as well, consistent with the spending behavior of the households involved. The results were then compared with the costs of the same items based

on average national prices without adjusting for regional differences in the cost of living.

The results are given in figure 1, which ranks all provinces (each one split into urban and rural areas) by the head-count index (that is, the percentage of people who are living below the poverty line) based on nominal expenditures, and then plots the results against corresponding estimates with an adjustment for differences in the cost of basic consumption needs. So, if the ranking of regions does not change when adjusting for the cost of living, the lines are everywhere positively sloped from left to right; if there is also agreement on the cardinal values of the poverty measures, the lines increase linearly on a 45-degree line (indicated by the straight line in figure 1).

It can be seen from figure 1 that both the poorest urban area (point A) and the poorest rural area (point B) are correctly identified. But in several cases adjustments for differences in the cost of living lead to a change in ranking. For example, the second poorest rural area after adjusting for cost of living differences (point C) is only the eleventh poorest without adjustment. The rank correlation between the head-count index using local poverty lines (with an adjust

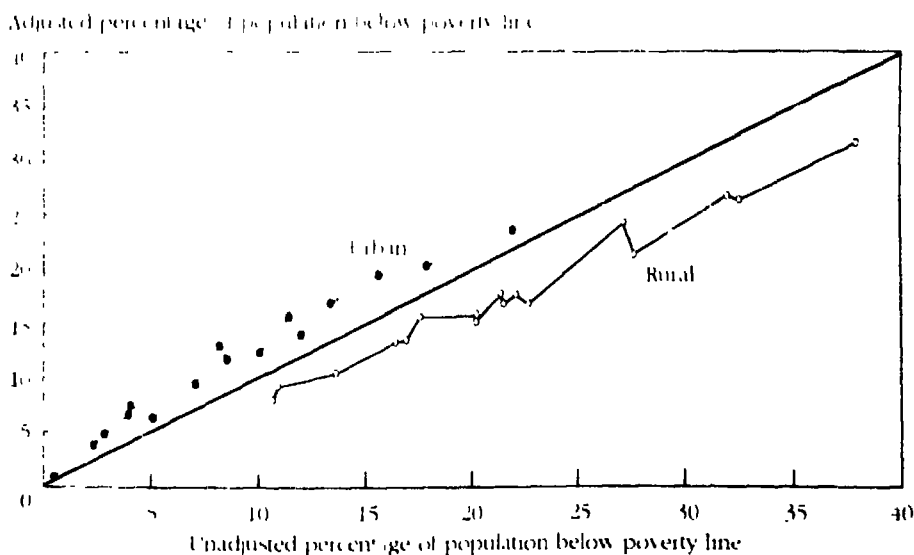
Figure 1. *Effect of Cost of Living Differences on Indonesia's Poverty Profile by Province*

Note: Each point is one province (urban or rural). The rank correlation between the adjusted and unadjusted figures is 0.88. Point A is the poorest urban area, point B is the poorest rural area, point C the second poorest rural area after adjustment for cost of living differences.

Source: Author's calculations using Indonesia's National Socioeconomic Survey for 1990.

ment for prices) and that using the national mean poverty line (that is, based on national mean prices) is 0.88.

If the experiment is repeated using sector of employment instead of location (urban or rural), price adjustment has much less effect because prices do not vary by employment (figure 2). Also note that for the sectoral profile, if differences in urban and rural costs of living can be incorporated correctly, the results with and without the adjustment for other regional differences show considerable agreement.

It is interesting to compare these results with another method that is commonly used to compensate for missing price data when setting poverty lines, the "food-energy-intake method." Indonesia and many other countries use this method, which defines the poverty line in each region as the level of expenditure at which a predetermined level of food-energy intake is typically met in that region. This technique has the advantage that it does not require any data on prices. But because people in richer regions or sectors tend to have more expensive tastes, and so spend more to reach any given caloric intake, the method can yield differences in poverty lines far in excess of the cost-of-living differences of the poor (Ravallion 1993). Ravallion and Bidani (1994) compared the food-energy intake method with a price-adjusted poverty profile for Indonesia and

Figure 2. *Effect of Cost of Living Differences on Indonesia's Poverty Profile by Employment*

Note: Each point is a subgroup of households defined by principal sector of employment.
Source: Author's calculations using Indonesia's National Socioeconomic Survey for 1990.

found that the differences in poverty lines between urban and rural sectors implied by the food-energy-intake method were large enough to cause a reversal in all poverty measures between urban and rural areas and numerous reversals in the ranking of provinces. After adjusting for differences in the cost of basic consumption needs, Ravallion and Bidani found a greater incidence, depth, and severity of poverty in rural areas than in urban areas. The food-energy-intake method indicated the reverse, however, with more poverty in urban areas. The ranking of regions (with each province divided into urban and rural areas) by the two methods was very different.

These inconsistencies arise because food-energy intake is determined by many other factors besides real incomes, such as relative prices, tastes, and activity levels, that can influence the poverty lines obtained by the food energy-intake method in ways that have nothing to do with differences in purchasing power over basic consumption needs. Group A may be able to afford more of all basic needs than Group B, and yet, because of the differences in the way the two groups allocate their money (say, group A buys more expensive calories because they taste better or are more convenient), the food energy-intake method can find that group A is poorer (Ravallion 1993). The method solves the problem of missing price data, but at a high cost in terms of the usefulness of the resulting poverty profile for informing policy decisions.

Some methods of dealing with missing price data work far better than others. And in some cases the best practice turns out to be the simplest. In the experiment above, not adjusting at all for regional differences in the cost of living gave better predictions of which regions or sectors were poorest than did the more sophisticated food energy-intake method. Even without data on prices by location, one can often make a reasonable guess as to how much more it would cost the rural poor to achieve the same standard of living in urban areas and to build this calculation into the poverty profile. This assumption can be based on a quick price survey that compares a few typical urban and rural areas.

Experiment 2: Monitoring Poverty Despite Gaps and Lags in Survey Data

There are often long lags between surveys, making it difficult to track a country's performance in fighting poverty. So developing a set of easily monitored, aggregate indicators of poverty that could serve as an alternative would be useful. An experiment in India compared the information provided by a set of aggregate indicators with survey data to see whether the findings from the two techniques were consistent.

This experiment used data supplied by India's National Sample Survey Organization, which has provided tabulations of the distribution of consumption per person since the mid 1950s. The surveys have been carried out at varying frequencies ranging from less than a year to more than five years. (For example, data were not collected between mid-1978 and 1983.) The consumption me-

sures are comprehensive, and the surveys appear to follow sound and consistent practices. The National Sample Survey (NSS) has been the main instrument for monitoring poverty in India, and a large body of literature uses these data (reviewed by Ravallion and Datt 1995). Analysts have had to rely almost exclusively on the survey agency's own tabulations of the raw data, which typically are not published until two to three years after the survey. Only rarely have the raw data been released, and in those cases the lags between collection and release appear to have been even longer.

Given these lags and the gaps between survey rounds, the question is whether poverty in India can be monitored using more readily available leading indicators from other sources. The answer would also help determine how frequently survey data should be collected.

Some simple forecasting experiments based on NSS data can try to answer the question. The analysis focuses solely on the rural poor, whose income is derived from their own farms (which often do not produce enough for their own food needs) and from wages received from working on other farms. Thus two potential leading indicators for monitoring poverty are the agricultural wage rate and agricultural output per acre. Both are typically available well in advance of tabulations from the NSS data.

Poverty measures, such as the head-count index, show that poverty increased in rural India during two periods—the early 1970s and the late 1980s (Ravallion and Datt 1995). How well could data on average wages and farm output have predicted these setbacks for the poor? The forecasting model used here to answer that question is based on a model of rural poverty outlined in detail in Ravallion and Datt (1995). The salient features of the model for the current discussion are summarized in box 1.

The model tracks the data well, but good within sample performance need not mean good post sample forecasts. To estimate the accuracy of this methodology as a predictor, two types of forecasts were considered.

- *Type 1.* This experiment entailed making forecasts one year ahead, assuming that the lag in the availability of NSS tabulations on the distribution of per capita consumption is one year longer than the lag in the availability of data on the real agricultural wage rate and agricultural output per acre. The extra lead time for these two variables would make it possible to predict poverty measures before the NSS data are released.
- *Type 2.* This experiment comprises a sequence of dynamic forecasts, again assuming that real wage and agricultural output data are available but that the NSS data are only available at longer intervals; one must thus rely partly on past forecasts of the poverty rate. (The previous year's poverty measure is one of the variables used to forecast the next year's poverty measure; see box 1.)

Table 1 gives the results of three forecasting exercises for the head-count index. (The results of forecasting two other poverty indexes—the poverty gap and

Box 1. *The Forecasting Model*

A separate model was estimated for each of three poverty measures; the head-count index, showing the percentage of the population below the poverty line; the poverty gap index, reflecting the average distance below the poverty line; and the squared poverty gap index, in which those distances are squared to make the measure sensitive to changes in inequality among the poor (Foster, Greer, and Thorbecke 1984).

The model was dynamic, in that the most recent past poverty measure was used as one of the predictors. Consumption-based poverty measures may not adjust instantaneously to current variables, so this feature is important. A nonlinear regression method was used to deal with the uneven spacing of the surveys over time (Ravallion and Datt 1995). The lagged poverty measure was a persistently significant predictor of the current measure.

The other two predictors were the real agricultural wage rate (deflated by the Consumer Price Index for Agricultural Laborers), and average agricultural output per acre in both the current and the last survey year. A time trend was also included to pick up the effects of any time-trended omitted variables. Although other variables were tried but had no additional explanatory power, it is clear that there are omitted variables. For example, from the data available it was not possible to construct a series for employment on rural public works schemes (which have historically been important in famine relief). Furthermore, lags and other limitations of existing public finance data appear to entail at least as large a lag as India's National Sample Survey (NSS) data, thus diminishing their usefulness as leading indicators. These forecasts could probably be improved, but it would entail considerable extra effort.

The model was first fitted to data for a core sample comprising all previous NSS rounds from 1958–59 until the date beyond which the forecasts were to be made. This entailed re-estimating the model for each experiment. Within-sample performance of the models was good; R^2 was 0.98 or higher when using the full sample of twenty surveys, and R^2 for the subsamples used for forecasting was in the range 0.90–0.98.

the squared poverty gap (see box 1) were not very different and are not reported.)

The results for type 1, in panel A, show forecasts of the poverty measures for 1987–88 (using a model based on the data up to and including 1986–87); for 1988–89 (using the data up to 1987–88); and for 1989–90 (using the data up to 1988–89). The percentage forecasting error ranges from a 7 percent underestimation to a 2.5 percent overestimation. The model failed to forecast the small increase in the head-count index between 1986–87 and 1987–88, predicting a drop instead.

Panel B gives type 2 forecasts for the same three years based on the same model as the 1987–88 forecast in panel A (using the seventeen survey rounds up to 1986–87). From 1987–88, new forecasts are drawn from the previously forecast poverty measure rather than from the actual measure. (This is the correct way to assess forecasting performance when there are gaps of more than a year between survey rounds.) Naturally, in this case, past forecasting errors are built into future forecasts, so a sizable drift of the forecasts from the actual values is possible relative to year-ahead forecasts. The first year in panel B is, of course

the same as in panel A. The second year is similar. But the third year drift considerably farther; the underestimation of poverty in 1988–89 magnifies th downward bias in the 1989–90 forecast.

We know that the living standards of the poor deteriorated in the early 1970 as a result of a drought in western India. How well does the model forecast thi event? Real agricultural wages *and* average farm yields fell in the early to mid 1970s. (See panel C for type 2 forecasts from the model using data up to 1970 71.) The model predicts the increase in poverty to 1973–74, but it overestimate the effect by a wide margin. Poverty did not increase nearly as much as th aggregate data on wage rates and yields suggest. The model correctly predict the end of the crisis, showing a decline in poverty by the late 1970s. By the nex survey round, however, it has drifted off track by a wide margin.

In summary, even highly aggregated data on agricultural wages and yiel contain information that is relevant to predicting poverty outcomes before ne household survey data are available or in lieu of them. But these are by n means perfect indicators. Although poverty measures can be forecast with tole able precision one year ahead of the release of the survey data, quite sizable dr can arise after just a year or two. In one case a model using these indicators fa to predict an increase in poverty, while in the other it does predict it, but su gests far worse outcomes than the eventual survey data. Possibly the forec could be improved by adding more variables, although the number of surve available and the lags and gaps in the availability of other data will no dou constrain such efforts in practice. It is likely that methods such as these [illegible] cheaper than conducting new surveys, but this saving must be weighed aga the loss of accuracy in monitoring poverty more than a year [illegible]

Experiment 3: Rapid Appraisal Methods

In a comprehensive socioeconomic survey, the bulk of the [illegible] devoted to various consumption and income components. For example, [illegible] cal survey for the World Bank's Living Standards Measurement Study [illegible] many pages of questions on consumption and income sources (Ainsworth [illegible] van der Gaag 1988; also see Demery and others [illegible] on [illegible] surveys [illegible] developed for African countries). Furthermore, calculating consumption [illegible] (particularly) income from the raw data is not easy. Some [illegible] argue [illegible] there are far better ways to assess living standards [illegible] wide range of [illegible] tions, including those described by Grootaert and Marchant [illegible], Kur (1993), Chambers (1994), and Narayan and Srinivasan [illegible]

During the last ten years or so, a technique known as "rapid appraisal" [illegible] emerged as a popular alternative to more comprehensive surveys based on st tistical sampling. The idea is to use very short interviews and direct observati to obtain a set of objective and subjective indicators of welfare. Some of th rapid appraisal methods include focus group and community interviews as w as less formal individual or household level interviews (see Kumar 1993). He

I focus solely on rapid household-level surveys, including subjective questions. Some exponents are convinced that rapid appraisal can be just as accurate as a full-blown survey (Chambers 1994)—an assertion that is difficult to verify unless both methods are applied to the same sample. Such an approach is rarely taken, however.

For this experiment, some typical rapid-appraisal questions were added to a high-quality conventional survey: the 1993 Jamaica Survey of Living Conditions, collected by the Statistical Institute of Jamaica. The survey already included many objective questions commonly covered in rapid-appraisal interviews; to these were added subjective questions on whether respondents considered the family's consumption of various categories of goods (food, housing, clothing, transport, health care, and schooling) to be "less than adequate," "more than adequate," or "just adequate." These questions, which required less than one page of the questionnaire and roughly five minutes of extra time, produced a response rate of virtually 100 percent. Forty-four percent of sampled households responded that their food consumption was "less than adequate for their family's needs," although the proportion rises to about two-thirds for the poorest quintile. A similar pattern in the differences between responses by "poor" and "rich" is evident from the answers to most of the other subjective questions (table 2).

How well can these subjective indicators predict a standard objective measure of consumption expenditure derived from the same survey? A regression model can be used to predict consumption per person, based on answers to subjective questions—several of which turn out to be highly significant predic-

Table 2. *Summary of Answers to the Subjective Welfare Questions for Jamaica*
(percent)

Answer	*All respondents*	*Poorest quintile*	*Richest quintile*
Food is inadequate	44.2	67.1	19.1
Housing is inadequate	48.7	71.7	26.2
Clothing is inadequate	40.6	64.1	16.3
Transportation is inadequate	51.5	65.4	34.0
Health care is inadequate	45.0	62.3	23.0
Schooling is inadequate	51.8	58.2	52.1
Food and housing are inadequate	32.5	57.2	10.4
Food, housing, and clothing are inadequate	25.7	48.9	6.8
Food, housing, clothing, and transport are inadequate	20.0	40.8	4.9
Food, housing, clothing, transport, and health care are inadequate	18.4	40.2	3.6
Food, housing, clothing, transport, health care, and schooling are inadequate	13.7	32.3	1.4

Note: Survey respondents were ranked according to household expenditure per person.
Source: Author's calculations based on the primary survey data from the 1993 Jamaica Survey of Living Conditions, collected by the Statistical Institute of Jamaica.

Overall, these experiments suggest that a rapid-appraisal survey lasting, say, twenty minutes on a relatively small sample can yield some significant predictors of the more expensive objective surveys. A significant share of the variance of consumption will almost certainly be left unexplained, however, and the predictions from such methods are unlikely to be highly correlated with actual standards of living. As with other short-cut methods described above, these limitations to the accuracy of rapid-appraisal methods will have to be weighed against the potential cost savings.

Experiment 4: Decomposing Aggregate Indicators without an Integrated Survey

Many socioeconomic indicators are available only in a highly aggregated form, that is, by provinces or countries. Yet one would like to know how they vary between socioeconomic groups, such as the poor and the nonpoor, or between urban and rural residents. This calculation would be easy if one had a survey that collected all the relevant information for the same households. Is it possible to estimate the decomposition reasonably well without such a survey?

Bidani and Ravallion (forthcoming) propose a method that can be used when the distribution of the population across the relevant subgroups (such as urban and rural) is known. An econometric model is estimated in which the observed aggregate indicator (such as the infant mortality rate) across countries or regions is modeled as a function of the population distribution by subgroups (such as the proportion of poor). Using this approach, the infant mortality rate of poor and nonpoor households can be estimated when all that is known is the overall average infant mortality rate for each country and the proportion of people who are poor. Bidani and Ravallion apply the method to find out how health indicators differ between the poor and the nonpoor. They find that the average life expectancy of the poorest two-thirds of the population in their sample of thirty-five countries is nine years less than that of the nonpoor and that their children are 50 percent more likely to die before their first birthday. Prescott and Jamison (1985), who use a similar technique to examine mortality rates and health service availability in China, found that rural residents had worse health and used health care less often than did urban residents. (The Prescott-Jamison method is not used here; for further discussion, see Bidani and Ravallion forthcoming.)

As with the earlier experiments, the most convincing way to test for bias in this method is through a counterfactual experiment. For this experiment, the accuracy of the urban-rural breakdown of provincial socioeconomic aggregates was tested. Many statistical agencies publish socioeconomic data at the level of the local administrative jurisdictions, but they rarely break down these figures into the categories one would like, such as urban and rural. Indeed, it may be administratively difficult to do so. Yet based on data in those countries that provide such information, large differences in socioeconomic outcomes, as well as in access to and use of public services, are known to exist between urban and rural areas.

How accurately can the urban-rural decomposition of socioeconomic aggregates be estimated? Five experiments were carried out for this article; four used data from provincial Indonesia, and one used data from the state level in India. In all cases, the aggregate provincial indicator, the actual urban and rural values of the indicator, and the shares of the urban and rural populations in the province or state were known.

The aggregate indicators were first related by a simple regression to the share of the population in urban and rural areas (using the "random coefficients" estimators described in Bidani and Ravallion forthcoming). The estimated coefficients represent the average urban and rural indicators across all the provinces. Table 4 gives the results for the actual and estimated decompositions of the means in each of the five cases. In only one case—the percentage of well-nourished children in Indonesia—are the estimates very close to the actual means. In the remaining cases, the rural mean is underestimated and the urban mean is overestimated.

Clearly, one of the problems with this test is that relevant factors that influence the socioeconomic indicator may have been omitted, thereby altering the model's estimates of the effect of the included variables. I therefore revised the model to account for some of these factors, as shown in table 4, with significant improvements in the estimated relationships. For food expenditure in Indonesia, the augmented regression used total consumption expenditure and average household size. For mean expenditure in Indonesia, the augmented regression used mean income. For income in Indonesia, the augmented regression used mean total expenditure. For the percent below a nutrition-based poverty line in India, the augmented regression used mean expenditure and the Gini index of expenditure. In the augmented models, I have deliberately aimed by trial and error to obtain the best estimates of the subgroup means using these explanatory variables from available data sets. In practice—when the true means are, of course, unknown—one would be unlikely to do as well as these estimates.

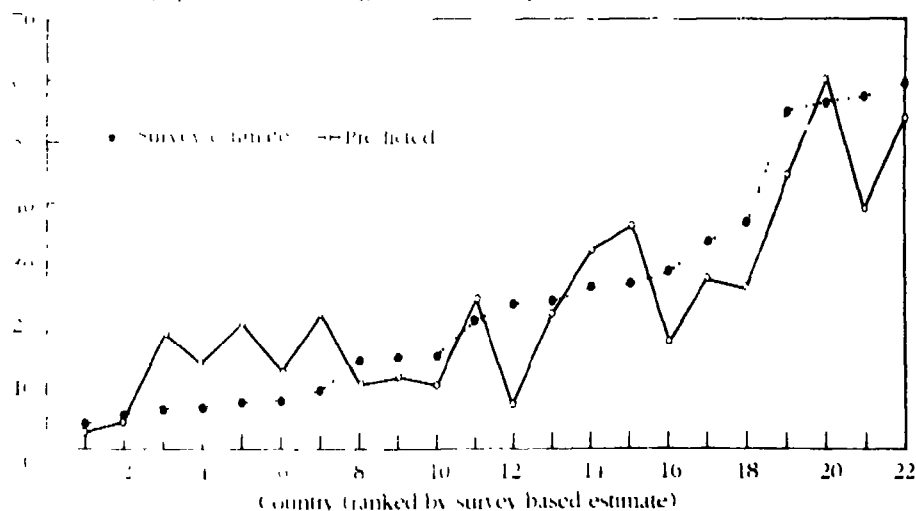
This experiment is an example in which the simplest methodology is likely to produce the least reliable results. The estimated urban and rural means are substantially closer to the actual means under the augmented regressions than under the simpler unaugmented regressions (table 4). Clearly, the urban-rural decompositions can be heavily biased unless the model is expanded to include a potentially wide range of other relevant variables. If augmentation can be done, the decomposition method is capable of yielding reasonable substitutes for the missing data.

Experiment 5: Assessing the Prevalence of Poverty without a Survey

This experiment was designed to show how accurately aggregate economic and social indicators estimate the prevalence of poverty. For this study, the proportion of people below the poverty line in each of twenty-two countries (estimated from household survey data) was compared with forecasts of the same

poverty measure in each country derived only from aggregate data. The poverty line was set at $1 a person a day (converted into local currencies to assure purchasing power parity in 1985). Using the methodology documented in Ravallion, Datt, and van de Walle (1991), regressions were estimated on each of the twenty-two possible samples, obtained by leaving out one country at a time. These are counterfactual experiments, meaning that the poverty measure for the particular country is assumed not to be known, although it is known for each of the other twenty-one countries. The predictor variables were private consumption per capita from the national accounts evaluated at purchasing power parity and at official exchange rates, the level of urbanization, the infant mortality rate, life expectancy at birth, and the proportion of women in the labor force. Some of these variables are of interest in their own right; infant mortality, for example, can be an important indicator of well-being (such as access to public health care), but here I am concerned only with the variable's usefulness in predicting consumption poverty.

Figure 3 shows that the predictions from the aggregate data are highly positively correlated with the survey-based estimates; the simple correlation coefficient is 0.87. Nonetheless, the absolute errors in predicting the prevalence of

Figure 3. *Actual and Predicted Poverty Measures across Countries*

Note: The predicted measure for each observation is based on a regression that excludes that observation.

Source: Author's calculations from primary data for twenty-two countries (Ravallion, Datt, and van de Walle 1991).

poverty are often quite large. Indeed, the average absolute error as a percentage of the original survey estimate was 49 percent. The poorest few countries were correctly identified, although there was considerable reranking. For example, the aggregate indicators suggested that the eleventh poorest country (based on the survey estimates) was almost the least poor. For the country with the largest absolute error, the survey estimate of the head-count index was 57 percent, while the aggregate economic and social indicators predicted a figure of 39 percent.

These discrepancies arise from two factors. First, the extent of inequality varies from country to country, and this variation is hard to pick up without distributional data from a survey. Second, the relation between social indicators and consumption-based poverty measures differs from country to country; some countries where a large share of the population is poor provide effective public health care and thus have good social indicators, such as low infant mortality, while others do not. These differences can make it hard to assess the extent of poverty without a household survey. It appears that the readily available economic and social aggregates can give, at best, a rough idea of the prevalence of poverty.

Conclusions

Much effort goes into making up for inadequate data on poverty; the tools used range from various rapid-appraisal methods to sophisticated econometric models. But in using such methods, how do we know that the cure is any better than the disease? Or how do we decide which cure works best? Surprisingly little effort has gone into rigorous testing of these methods. New approaches—aiming to implement some new methodological paradigm or just to improve an existing survey—are routinely introduced without prior testing. Little more than blind faith guides the policymaker's interpretation of results.

In many ways the experiments reported here represent a limited investigation. It is not feasible to span all the methodological choices that matter, although these experiments may be suggestive in other applications. The results may not be robust to measurement error or applicable in other settings; there are still precious few opportunities for tests of this sort. And there are gains and losses that I have not been able to quantify. The imprecision introduced by imperfect data and methods may appreciably lower the expected benefits from efforts at targeting the poor. Nor have I addressed the costs of data collection, which must be weighed against the benefits of greater accuracy in measurement.

Nonetheless, the results do suggest that there are some relatively easily monitored proxies for poverty that can help identify and monitor poverty when survey data are unavailable or inadequate. These proxies rely on certain plausible theoretical or common sense relationships that are expected to hold between

conventional survey-based poverty data and more aggregated data from other sources (or from simpler rapid-appraisal methods). The experiments reported here suggest that data inadequacies can be at least partially surmounted by careful use of such methods (table 5). Even with rather poor data, the situation facing a creative analyst is rarely hopeless.

The results also point to the limitations of these methods. Omitted variables may well account for such a large share of the variation in standards of living that the method tends to be a poor substitute for better data. Even a rather narrowly defined objective indicator of welfare, such as consumption or income, is determined by a large number of variables, only a small subset of which are amenable to monitoring by these methods. In addition, the variation in welfare at any given income level is so broad that one must remain somewhat pessimistic about the scope for really credible monitoring of poverty in many data-poor settings.

Table 5. *Summary of Experiments and Results*

Data problems	*Proposed solutions*	*The tests*	*Results*
1. You have a good consumption survey, but no locational cost of living index.	Either (1) ignore the problem and use one nominal poverty line in all regions, or (2) anchor the lines to food-energy requirements so that the local line is the expenditure level at which requirements are met on average in each region	First estimate the regional poverty profile for Indonesia using locational data on the costs of basic consumption needs, then see how well the profile can be predicted without these data.	Method (1) is better than (2). The errors in ranking using (1) need not be large, particularly if one can at least make a reasonable assumption for the urban-rural cost-of-living difference. But method (2) can be way off the mark when aiming to construct a profile of absolute poverty.
2. There are gaps and lags in survey data availability	Identify proximate causes of poverty that can be monitored using more readily available aggregate data	Using time series data for rural India, see how well the model predicts beyond the period over which it was estimated. Use real agricultural wage rate and average farm yield as the predictors	The wage rate and farm yield are good predictors of rural poverty measures for India. One can forecast poverty fairly well one year ahead of the survey. But sizable drift can arise after one year
3. A full-size objective socioeconomic survey is not feasible, but a rapid appraisal is	Use soft survey [illegible] methods, [illegible] questionnaires [illegible] simple [illegible]	A [illegible] sized conventional survey for [illegible] included a short module of subjective and simple [illegible] questions, so the results can be compared [illegible] for consumption	Roughly half of the variance in consumption can be explained this way. But the method still entails large errors; for example, in 60 percent of cases the rapid-appraisal method incorrectly predicts the decile of those actually in the poorest decile. The method was better at identifying rich households than poor ones

Notes

Martin Ravallion is with the Policy Research Department of the World Bank. He would like to thank Benu Bidani, Shaohua Chen, Gaurav Datt, Christiaan Grootaert, Margaret Grosh, Stephen Howes, Emmanuel Jimenez, Dominique van de Walle, Jacques van der Gaag, and Qing-hua Zhao. Statistical agencies in twenty three countries have helped by collecting data used here, but special mention should be made of the Statistical Institute of Jamaica for agreeing to add to its survey the questions used in Experiment 3.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems

Ainsworth, Martha, and Jacques van der Gaag. 1988. *Guidelines for Adapting the LSMS Living Standards Questionnaires to Local Conditions*. Living Standards Measurement Study Working Paper 34. Washington, D.C.: World Bank.

Bidani, Benu, and Martin Ravallion. Forthcoming. "Decomposing Social Indicators Using Distributional Data." *Journal of Econometrics*.

BPS (Biro Pusat Statistik). 1989. *Welfare Indicators*. Jakarta.

Casley, Dennis J. 1993. "Introduction." In Krishna Kumar, ed., *Rapid Appraisal Methods*. A World Bank Regional and Sectoral Study. Washington, D.C.: World Bank.

Chambers, Robert. 1994. "The Origins and Practice of Participatory Rural Appraisal." *World Development* 22(7):953–70.

Chaudhuri, Shubham, and Martin Ravallion. 1994. "How Well Do Static Welfare Indicators Identify the Chronically Poor?" *Journal of Public Economics* 53(3):367–94.

Cowell, Frank, and Stephen Jenkins. 1994. "How Much Inequality Can We Explain? A Methodology and an Application to the USA." Distributional Analysis Research Programme Working Paper 7. London School of Economics, London. Processed.

Datt, Gaurav, and Martin Ravallion. 1995. "Regional Disparities, Targeting and Poverty in India." In Michael Lipton and Jacques van der Gaag, eds., *Including the Poor*. Washington, D.C.: World Bank.

Demery, Lionel, Jean-Luc Dubois, Christiaan Grootaert, and Tim Marchant. 1992. "Annotated Questionnaire." In Ghislaine Delaine and others, eds., *The Social Dimensions of Adjustment Integrated Survey: A Survey to Measure Poverty and Understand the Effects of Policy Change on Households*. Social Dimensions of Adjustment in Sub-Saharan Africa Working Paper 14. Washington, D.C.: World Bank.

Foster, James, J. Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52:761–65.

Grootaert, Christiaan, and Timothy Marchant. 1991. *The Social Dimensions of Adjustment Priority Survey: An Instrument for the Rapid Identification and Monitoring of Policy Target Groups*. Social Dimensions of Adjustment in Sub-Saharan Africa Working Paper 12. Washington, D.C.: World Bank.

Haddad, Lawrence, and Ravi Kanbur. 1990. "How Serious Is the Neglect of Intrahousehold Inequality?" *Economic Journal* 100:866–81.

Kumar, Krishna. 1993. "An Overview of Rapid Appraisal Methods in Development Settings." In Krishna Kumar, ed., *Rapid Appraisal Methods*. A World Bank Regional and Sectoral Study. Washington, D.C.: World Bank.

Narayan, Deepa, and Lyra Srinivasan. 1994. *Participatory Development Tool Kit: Training Materials for Agencies and Communities*. Washington, D.C.: World Bank.

Prescott, Nicholas, and Dean Jamison. 1985. "The Distribution and Impact of Health Resource Availability in China." *International Journal of Health Planning and Management* 1:45–56.

Ravallion, Martin. 1993. *Poverty Comparisons.* Fundamentals in Pure and Applied Economics, vol. 56. Chur, Switzerland: Harwood Academic Press.

Ravallion, Martin, and Benu Bidani. 1994. "How Robust Is a Poverty Profile?" *World Bank Economic Review* 8(1):75–102.

Ravallion, Martin, and Gaurav Datt. 1995. "Growth, Wages, and Poverty: Time Series Evidence for Rural India." Policy Research Department, World Bank, Washington, D.C. Processed.

Ravallion, Martin, Gaurav Datt, and Dominique van de Walle. 1991. "Quantifying Absolute Poverty in the Developing World." *Review of Income and Wealth* 37(December):345–61.

Ravallion, Martin, and Binayak Sen. 1994. "Impacts on Rural Poverty of Land-Based Targeting: Further Results for Bangladesh." *World Development* 22(6):823–38.

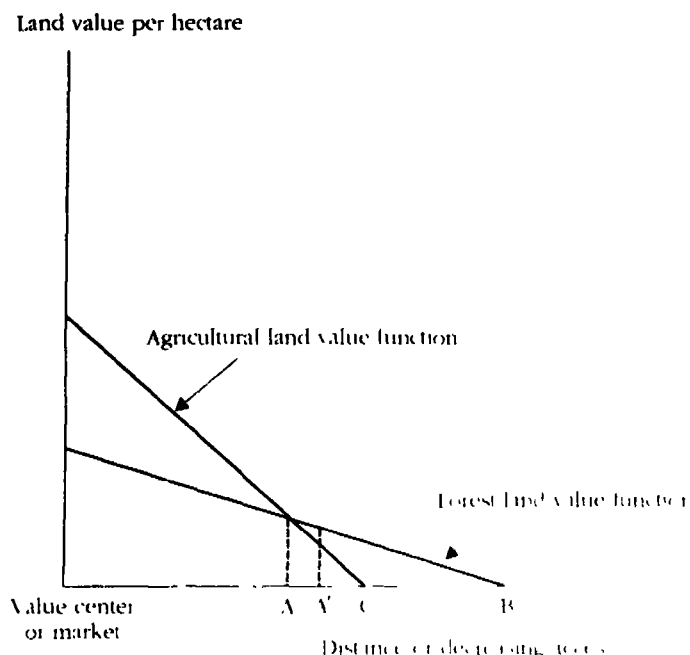
DEFORESTATION AND FOREST LAND USE: THEORY, EVIDENCE, AND POLICY IMPLICATIONS

William F. Hyde
Gregory S. Amacher
William Magrath

The topic of deforestation is seldom examined from the perspective of prices and responses to resource scarcity. This omission creates important errors in policy. Resource scarcity induces investments in both commercial and subsistence uses of the forest once prices overcome the costs of establishing property rights, forest management, and the returns from alternative agricultural uses of the land. Therefore deforestation will induce price increases and investments in forestry well before deforestation attains its physical limit. These prices and costs will alter the boundaries among several important classes of forest land: sustainable private forestry, the forested commons, unsustainable open access forests, and unused residual forest. The greatest impact on the world's forests will come from refocusing the policy dialogue on the cost factors that determine these boundaries, including agricultural support policies, local concentrations of nonmarket environmental resources, and policy failures that distort incentives to invest in forestry. In locations where reforestation induces large price changes, policymakers must remain attuned to the likelihood that deforestation induced changes in the prices of forest products and forest policies may cause significant shifts in the activities of the poorest people.

Conserving scarce forest resources is a challenge for both high- and low-income countries. Contemporary forest policy and forest management both reflect this challenge. Forest policy reflects international concern with the pressures of deforestation, including trade in tropical timber, the conversion of forests to agricultural uses, and the effects of deforestation on climate change, biodiversity, and local communities dependent on forest resources. Forest management features the transition to managed forests as deforestation causes

The World Bank Research Observer, vol. 11, no. 2 (August 1996), pp. 223–48

Figure 1. *The Relationship between Land Value and Market Access*

Note: Point A is the point at which agricultural land value equals forest land [illegible] point B is the point at which short run forest land value equals zero [illegible] the point where agricultural land value declines to zero [illegible] no private value

and communities will establish clear rights to this part of the forest if the long run value exceeds the costs of establishing and protecting the property rights. Somewhere between points A and B, the costs of establishing and protecting permanent rights to the land exceed the in situ value of any products harvested from the forest, and it becomes an open-access resource. Thereafter, the forest value function in figure 1 reflects only short run or extractive resource values. Generally, the costs of establishing and protecting permanent rights to the land are greater than the value of the land for long run forest production for all land beyond point A (agricultural uses reclaim the land up to point C, and all of the forest between points C and B is an open-access resource).

This is an important point. It suggests why observed rates of harvest, deforestation, and land conversion are not as socially excessive as they sometimes seem and also why establishing property rights is not an easy solution. Figure 1 shows that the values of standing forest resources are low relative to other land uses. Forest resources may also be dispersed. For example, fruit and nut-bearing trees

tend to be scattered, and some high-value tropical timber species occur only on an occasional hectare. The costs associated with establishing and maintaining property rights for these resources can easily exceed their in situ values. Fences, forest guards, and roadblocks may help protect these rights, but trespass and theft are common even in the forests of industrial countries. Inevitably, at some point between A and B in figure 1, the value of the forest will be positive but less than the cost of secure property rights. Open-access forestry is efficient in this region.

Finally, the standing forest located beyond point B is also an open-access resource, but extraction from this forest is not economically attractive. This forest is usually the stewardship responsibility of the forest ministry or national park agency. Many of the forests of interior Alaska, northern Canada, Siberia, and some of the Rocky Mountains in the United States, and most remaining tropical forests, including much of the Amazon, fall into this category. This economically residual resource accounts for a significant part of most recorded measures of global forest inventories despite its uneconomic nature because the universal convention for forest inventories is a physical measure taken without regard for access costs. Therefore, access and harvest costs for industrial users and the opportunity costs of labor for subsistence forest collectors create very real limits on the extraction of resources, and the costs of ensuring tenure create a very real burden for alternative strategies for resource management.

These arguments and the description shown in figure 1 accommodate many nonmarket forest values (as well as the obvious commercial forest values. Fuelwood, forage, fodder, fruits, and nuts, for example, actually do exchange in local markets, and subsistence users of the forest choose the levels of their market participation according to the opportunity costs of the labor used in collecting the fuelwood, forage, and so on. Therefore, the costs of deforestation, land conversion, and property rights do not reflect the magnitude of the losses that would be implied by an absence of markets. Nevertheless, some nonmarket environmental values do remain (notably watershed protection or erosion control, biodiversity or genetic reserves, and carbon sequestration to protect against global climate change), and these will be discussed later.

Scarcity and Increasing Prices

Deforestation, climate change, and biodiversity are largely concerned with the deterioration of the residual stock of natural forest to the right of point B in figure 1. This is a dynamic problem. It is affected by policy changes, and it will change as resource extraction continues over time. It is reasonable to expect that extraction from the natural forests will continue until the prices of forest products attain a level that induces investments in trees or forests. The key issue for policy formulation should be the relative magnitudes of eventual prices for forest products, the eventual location of active tree or forest management, and

the final combined inventory of planted trees and natural forests—or the final status of world deforestation.

In brief, as the extraction of forest resources continues, the margin of the natural forest (point B in figure 1) must shift outward (to the right), forcing the residual natural forest to decline. If society continues to consume forest resources, then extraction must continue; access and harvest costs, including the opportunity costs of subsistence labor, must rise; and the market price of forest products must rise with them. As the stock of natural forest declines and access costs and market prices continue to increase, the function representing forest resource value in figure 1 rises until market prices (or the opportunity costs of subsistance households) eventually justify the costs of investments in trees and of securing and protecting their property rights. If these combined costs are less than the price represented by the forest land value gradient at some point to the right of A (say A′), new plantations will appear between that point and point A to its left.

Land Tenure Rights

Agricultural land extends to point A, where the agricultural gains are equal to the costs of securing agricultural property rights. If the cost of securing the rights to forest property is comparable to the cost of securing agricultural rights—which is the general case—a discontinuity in the use of forest land can be anticipated. Once the initial harvest of the natural forest between points A and B has been completed, agricultural use extends to point C. The extraction of forest products continues at the ever more-distant margin (point B shifts to the right) and the values of market and subsistence forest-based goods continue to increase. Eventually prices of forest products will overcome the costs of securing rights to forest property and the costs of local reforestation and forest management. Once these prices rise further to the level of the agricultural opportunity, trees will compete as an agricultural crop in the neighborhood of point A. This pricing sequence argues that trees will become an efficient activity somewhere closer to the market center than some extensive agricultural activities.

The increasing recognition that private investment responds to declining natural forest stocks and increasing prices for forest resources is usually associated with fuelwood and other subsistence forest products (Hofstad 199[illegible]; Mercer and Soussan 1992; Cline-Coal, Main, and Nichol 1990; Dewees 1989). Farmers commonly plant trees near their households or in rows along well-traveled fences, paths, and nearby embankments to provide better protection for their higher-value forest resources. (Regular surveillance is more difficult and theft is easier at the more distant fringe of the household's agricultural lands.) Rural households follow this strategy in developing countries around the world. Their private plantings may consist of only a few trees (not enough to be called forest "plantations"), but in a few cases these private plantings have been sufficient to replace all the resources removed from the open-access natural forest.

Commercial forest plantations in industrial countries also tend to follow this model. Although they do not compete with the highest agricultural values, they often produce a more valuable crop than agricultural uses such as grazing. The land-use sequence (see figure 1) ranges from high-value agriculture to commercial forest plantations to livestock grazing to timber harvests (from the remaining natural forests) to an unharvested residual forest. Sedjo and Lyon (1990), for example, project that this pattern will prevail for the foreseeable future, with at least half the world's harvest of industrial wood in 2050 originating from natural forests. Therefore, periodic harvests from natural forests will play a large role in industrial timber supply beyond 2050.

Classification of Forest Land

The investment response to the increasing scarcity of forest resources creates four or possibly five categories of forest land. The first is an economically residual standing natural forest to the right of point B in figure 1. The second is a region of mature natural forest from which timber or other forest products are currently harvested (the neighborhood of point B itself). The land area and forest inventory in this second category are highly responsive to short-run changes in prices, harvest and access costs, or forest sector policies. Higher prices for forest products, technical changes that reduce extraction costs, new roads into previously inaccessible regions, and the reduction of public policy constraints on harvesting each expand the area and volume of current resource extraction. Declining markets and tighter policy constraints reduce the size of this category of forest land and decrease or delay current extraction. These two categories of land use are the common focus of most forest policy discussions.

The third category of forest land covers the region of previous resource extraction to the left of point B. It supports repeat or pulse harvests as the depleted natural forest grows to some minimum economic size. Access to the resources in this region remains open because the cost of obtaining and protecting property rights exceeds the potential return to forest investment.

A fourth category of forest land may appear in the neighborhood of the extensive agriculture (to the right of point A) if the expected in situ return on forest investments exceeds the costs of protecting rights to forest property. This land use category is more likely in subsistence communities where common ownership offers a lower cost and a more secure regime of property rights than do individual private rights. The initial cost of converting forests to agriculture identifies the boundary between this fourth category of forest land and extensive agricultural uses of land.

This category cannot exist when the cost of secure property rights exceeds the expected return on investment in forest resources. The prices and costs of forest products continue rising until increasing demand eventually shifts some forest production for both subsistence and industrial uses to a fifth category of higher-valued (previously agricultural) land. The in situ forest product prices for

this land must justify the cost of trees managed as an agricultural crop, includ ing the important cost of secure property rights. This land-use category is a recent phenomenon in forest history that was unnecessary before deforestation induced scarcity.[2]

These observations suggest that three categories of forestry are sustainable commercial plantations, household tree plantings, and the uneconomic residua natural forest. The regions of current resource extraction and pulse harvests wil always be subject to open-access extraction because the cost of securing tenur in these regions exceeds the value of the resource. These regions can becom sustainable only with effective, and perhaps expensive, public regulation or for estry ministry management. That expense may be unacceptable in many devel oping countries, and the degradation of forests in these regions may be inevi table and economically efficient.

The Empirical Evidence

Data on these five categories of land use are exceedingly difficult to obtain Table 1 gives a rough impression of the importance of each category in a cross section of five countries. These data were collected in a variety of formats. None were collected with the intention of identifying the economic distinctions of land use, and they seldom provide the means to estimate the areas of pulse harvests. These areas of degraded forests no longer appear in the forest statistics of many countries. Nevertheless, a few points are clear. Plantations and trees planted for household uses account for significant areas, even in some develop ing countries, and the areas of uneconomic residual forest are extensive in the five sample countries. The evidence on nontimber products is limited; the evi dence in table 1 refers to harvests of commercial timber. The margin of natu forests is an important source for commercial harvests in the United States. I

Table 1. *Classification of Forested Regions*

thousands of hectares

Country	Commercial and household plantations	[illegible]	[illegible] harvests from natural forest	[illegible]
Belize	3		[illegible]	[illegible]
Philippines	290		[illegible]	[illegible]
Chile	1,400		Negligible	[illegible]
United States	2,000		[illegible]	[illegible]
Finland	12,650	[illegible],000		[illegible]

- - Not available

Source: Derived from Chomitz and Gray (1994), Republic of the Philippines (198[illegible]), Hyde and C[illegible]berlain (1995), U.S. Department of Agriculture (1992), Aarne (199[illegible]), Banes (199[illegible]), World Reso[illegible] Institute (1994)

the site of agricultural land conversion, if not for commercial timber, in Belize and the Philippines.

The evidence in table 1 is static. The real support for this argument depends on changes in land use and the increasing scarcity of forest resources over time. Scattered evidence on rapidly changing patterns in land use for both industrial and developing countries supports the contention that scarcity does generate investments in forestry.

Industrial Forestry: The United States

There is no reliable long-term record of stumpage prices (the value of standing timber) in the United States, but lumber prices have risen at an inflation-corrected annual rate of 1.8 percent since 1870 and perhaps longer. This is the only long-term real price increase in U.S. history for any primary natural resource (Barnett and Morse 1963; Ruttan and Callahan 1962). Of course, prices cannot continue to climb indefinitely; Barnett and Morse (1963) and Berck (1979) predicted that the rate of increase would decline, and more recently Sedjo and Lyon (1990) contended that the upward movement came to an end during the past twenty years.

As the stock of natural forests was drawn down during this long period, its depletion drew the attention of the American Association for the Advancement of Science in 1873 and provided the justification for the Forest Reserve Act of 1897, which became the basis for the National Forest System. The U.S. Forest Service, in approximately decennial projections from 1909 to the present, continues to anticipate timber shortfalls, but Clawson (1979) pointed out that Forest Service projections have never been correct, largely because they always underestimate future production and price induced improvements in wood use and harvest technologies.

Libecap and Johnson (1979) observed that settlers of the forested American frontier harvested the forests and claimed the land as soon as the land had value—to the point at which all value was dissipated in the costs of acquiring the property rights. Furthermore, in separate analyses, Johnson and Libecap (1980) and Berck (1979) observed that fears of excessive harvesting from the natural forests in the Great Lakes region from 1880 to 1920 and from the Pacific Northwest since 1950 were unwarranted. Indeed, for these two great timber-producing regions, real stumpage prices rose at less than the rate of interest, and harvests of the natural stock occurred at rates considerably below expected returns on private capital. After a long period of apparently rational private extraction, the accompanying price increases have slackened, forest investments have become profitable, and the modern U.S. forest industry supports more than 5 million hectares of private plantations.[3]

Industrial Forestry: Other Countries

Chile is another industrial wood basket of the late twentieth century, and it provides a similar story—although without the supporting econometric litera-

ture. In the nineteenth century, the demand for mine timbers led to extensive depletion of the country's natural forests. The large mining interests responded by planting experimental forests to supply their future needs. Agricultural cultivation kept the natural forest in abeyance, and pulse harvesting of the natural forest growth largely satisfied the remaining domestic demand until the 1950s. Since then, expanding demand from Europe, the development of Chile's ports, and (arguably) a government program of forestry assistance, have provided a supporting environment for more than 1.4 million hectares of forest plantations. These plantations are generally located at the margin of good agricultural land; but within much easier access to the ports and processing facilities than the natural forests (Vincent and Binkley 1992; Amacher and others 1996).

Several developing countries have begun the transition to plantation production of industrial wood. In Costa Rica the conversion of forests to agricultural land and the growth of industrial and subsistence forest consumption have contributed to deforestation. Little of the original forest remains, and the remaining natural forest is insufficient to ensure sustained operation of the local wood processing industries. One firm, recognizing the potential long-run gains to continued operation of its mills, has invested in natural forest management on 10,000 hectares in an area in which it has secure property rights (Rice 1993).

In Kenya 90 percent of the industrial wood came from the natural forest in the 1950s. Today 80 percent comes from 180,000 hectares of pine and cypress plantations, and less than 10 percent is harvested from natural forests (World Bank 1992). Although it is clear that land conversion for agriculture is the primary source of deforestation, it is reasonable to assume that forestry only recently began to offer the satisfying financial return necessary to induce private investment. Scherr's (1995) historical evidence for western Kenya paints a clear picture of scarcity-induced household investments in trees following permanent settlement in that region early in this century.

In Malawi, tobacco farmers are the largest consumers of industrial wood both for posts and for fuel in flue curing. They relied on natural forests on customary lands (lands largely open to first claimants) until Malawi's 3.5 percent annual rate of deforestation and more than 5 percent increase in the price of fuelwood induced both a new wood-saving technology for curing tobacco and self-sufficiency in wood production. Some of the new forest plantations are in large blocks in less populous regions where forest trespass is not a serious problem. Other tree plantings, in more populated regions where trespass is more likely and the protection of forest property rights is more difficult, tend to be in fence rows, farm yards, and along paths and roads near farm operations (Hyde and Seve 1993).

In the Mekong delta of Vietnam, small farmers have responded to the high export price for wood chips (for pulping) by introducing a new tree species, *eucalyptus camalduensis*, for sale to the new export market. They cultivate the new species in private stands and along paths and paddy edges, but always close to the household where private investments are most secure (Byron 1993).

Forest Production for Household Consumption

Farmers in many developing countries have responded to higher prices by planting trees to satisfy household consumption. Several examples from Nepal serve as nice illustrations, in part because Eckholm (1976) made the world so aware of deforestation in that country. Most of the evidence is cross-sectional and is restricted to fuelwood, although many other forest resources (fruit, nuts, forage, and fodder) consumed by subsistence households also exchange in local markets. The patterns of forest extraction and investment for these resources are probably similar to those observed for fuelwood.

Shifting cultivators receive much of the blame for Nepal's deforestation, but careful observations show that they respond to increasing scarcity by settling, establishing common and then private property rights and ultimately introducing more land-intensive and longer-term agricultural technologies (Eggertsson 1990 reviews this literature). The Gurung population of the Annapurna Sanctuary followed precisely this pattern during the course of this century (Stevens 1988). Early in the century this region was isolated, even for Nepal, and the Gurungs cut the forest, grew subsistence crops for a few years, and moved on to cut more forest. Over time population pressures, deforestation, declining agricultural yields, and growing familiarity with new agricultural technologies induced major social changes. The indigenous population is now settled. Private property rights are well defined, and the Gurungs use agricultural technologies such as terracing that are widespread throughout Nepal. Some Gurung landowners plant trees, and the remaining forest has stabilized.

Gautam and others (forthcoming) surveyed small landowners in five communities in Nepal's hills. Many farmers own trees and use them for fuel and construction, as a source of fruit and fodder, and for other purposes. Their private plantings of trees have increased in the past ten years as the forest inventory on common lands has declined. More often, it was the wealthier and better educated landowners who planted trees and adopted other conservation techniques.

Amacher, Hyde, and Joshee (1993) examined household production of fuelwood in two hill districts of Nepal. Fuelwood prices, distance from common forest lands, and the standing forest inventory differed significantly in the two districts. Not all households bought or sold fuelwood, but many did, and all had the opportunity to participate in local markets. Households in the district characterized by lower prices, briefer travel time, and a larger forest inventory (all indicating that scarcity is less serious) relied on fuelwood from the forested commons. Households in the district where fuelwood was scarcer produced more fuelwood on their own lands. Scarcity was also more likely to induce households in the latter district to substitute combustible agricultural residues for fuelwood and to adopt improved cook stoves. A broader survey of households in twenty-nine (of fifty-nine) districts in Nepal's more populated hill and tarai regions produced econometric production and consumption results consistent with the first analysis and with the hypothesis that subsistence households respond to

increasing resource scarcity by increasing their plantings of trees on private land (Amacher, Hyde, and Kanel forthcoming).

In another study on Nepal's tarai, Kanel (1993) reported annual price increases of 16 percent for fuelwood and more than 20 percent for construction wood. He also reported a 1.3 percent rate of annual deforestation regionwide and total deforestation of the region's public forestlands. Private plantations totaled 10,000 hectares in 1981 but expanded to 150,000 hectares by 1991. Barnes and others (1993) made similar observations for the more populated periurban areas of Java, and Krutilla, Hyde, and Barnes (1995) drew similar conclusions for an even broader collection of thirty-three periurban areas in Africa, Asia, the Middle East, the Caribbean, and Latin America.

The counterargument to scarcity-induced investment is that some species and associated forest resources grow only in natural environments. These species are unresponsive to scientific intervention and cannot be domesticated. Brazil nuts are the common example, but Viana and others (1996) have shown that deforestation and higher prices induce successful plantations even in this extreme case. Industrial examples of rubber, oil palm, and coffee trees, and subsistence economy examples of many fuelwood, forage, and fodder species are only better-known cases of the same effect.

Nonmarket Values and Resource Degradation

The argument that market-induced responses to increasing scarcity will correct many forest allocation problems is appropriate for commercial timber as well as for many nontimber forest resources. It leaves unaddressed three issues of great importance: erosion control, biodiversity or genetic reserves, and the control of global climate change.

Controlling Erosion

Erosion is a local problem caused by livestock, shifting cultivation, and logging activities on forested uplands. It affects the adjacent lowlands and watercourses within a watershed. Its solutions must also be local. One solution arises with the shift from forest extraction at the economic fringe to plantation management in higher-value lowland areas. This shifts the eroding activity away from more fragile upland environments and also provides tree cover for more heavily used lowlands. Both steps lessen erosion, but the problem will remain in many areas. Two empirical analyses demonstrate the potential for regional economic gain. Anderson (1987) estimated the expected effects of reforestation on wind control and increased agricultural productivity in northern Nigeria. Yin and Hyde (1995) calculated an 8 percent gain in agricultural productivity attributable to increased tree cover in northern China since 1974.

The evidence from industrial countries (for example, McConnell 1983 and Miranowski 1984) argues that individual landowners correct for the problems of erosion when they own both the sources and the effects of soil loss. This finding suggests that any system that improves private rights for forested property and enables communal arrangements for addressing transboundary problems would also decrease erosion. Where improved property rights and community-imposed landowner rules are insufficient, reforestation and broader restrictions on human access to the forested uplands may be necessary. Enforceable property rights are a prerequisite, however, for effective broader controls and successful public interventions. Their absence may justify custodial responsibility by the forest ministry, but effective administration of custodial responsibilities is an uneasy challenge. Few forest ministries have shown that they are up to it.

Preserving Biodiversity

Biodiversity and genetic reserves are broad international values conferred by highly specialized and generally local forest resources. There is some evidence of a rising international market for the protection of genetic resources (Simpson, Sedjo, and Reid, forthcoming) and some evidence that these values are not large in any aggregate sense (Sedjo 1992; Mendelsohn and Balick 1995). In any event, the market is thin, and these resource values often have to be protected with specialized management criteria or in specialized local forest reserves. The problem is to arrange secure rights for specialized habitats.

Protecting biological diversity is all the more difficult because neither flora nor fauna respect property boundaries. Moreover, protecting a few select individuals or even an ecosystem does not guarantee either diversity within a species or the species' long term survival. Protecting biodiversity requires land-use arrangements that are more complex than fences and permanent restrictions on forest management in national parks and wildlife preserves.

Furthermore, the protective land-use arrangements must change depending on the type of forest. In tropical forests, local endemism and species specialization are greater. The key issue is ensuring undisturbed regeneration cycles in individual, site specific populations. In boreal forests the key issue is a shifting forest mosaic with respect to species and age structures. Species turnover or progression is more rapid in boreal forests. Therefore, tropical forests may contain more species diversity, but boreal forests may require larger areas to protect a single species (Vehkamaki and Simula 1995).

As figure 1 shows, the critical forested regions for both erosion and genetic reserves are the open-access forest between points C and B and the region of extraction from mature forests in the neighborhood of point B. The residual natural forest beyond point B is undisturbed by forestry activities; its soil and genetic reserves also remain undisturbed. Forest plantations, in the region of point A, protect the soil, but this region long ago lost most of

its contributions to natural genetic diversity. It is in extensive agriculture (A-C) or the open-access forest (C-B) and the exploitable mature natural forests (the neighborhood of B) where local enclaves of nonmarket environmental values such as watershed protection (erosion) and genetic diversity justify public policy interventions.

Global Climate Change

The importance of global climate change is an unresolved issue, but, if it is important, there can be no doubt that trees provide an important carbon reserve and protection against further global change. All trees store carbon. The storage of carbon increases as trees are planted, grown, cut, stored, and new trees planted. Incorporating a value for protection against global climate change enhances the value of standing forests. In terms of figure 1, it raises the forest value functions and shifts point A to the left, which means that it would have been socially optimal for commercial forest to compete successfully with agriculture at lower forest product prices and at an earlier time. It expands the socially optimal area of land in commercial forests. The social valuation of carbon sequestration, however, is difficult to calculate and even more difficult to impose on land management because this value is held by many consumers who have little personal contact with the forest itself. Furthermore, policies addressing global climate change would also have significant administrative costs and high costs per tree or hectare protected. Therefore, one might hypothesize that correcting those policies in other sectors that have large deleterious spillover effects to forestry and improving the security of property rights to the forest may cause greater shifts in land use at points A and B and save more trees for carbon sequestration than would altering prices and policies to reflect optimal adjustments for global change.

Policy Implications

The first conclusion of a model that shows responses to rising prices is that deforestation is not the ultimate issue often portrayed in popular discussion. Rather, the market sets a limit on the extent of deforestation once prices rise to a level equal to the "backstop" costs of plantation forestry (that is, the cost of planting and managing trees or forests plus the cost of ensuring the rights to the investment).

Important delays may precede the initial investment in tree planting. The first delay occurs while the natural forest is drawn down to the point where its harvest and access costs equal the plantation backstop price. Investor uncertainty due to the long production cycles typical for forests may cause an additional delay. These delays could be arguments for policy intervention, but there is no empirical evidence of their economic importance. One problem is the difficult

of measuring the aggregate scarcity response of many small landowners. Byron (1984), for example, reported that three-quarters of all fuelwood consumed in Bangladesh comes from forest stands that are too small to be included in the national forest inventory.

Sedjo and Lyon (1990) found that the economic limit on deforestation has been attained for industrial timber demands. Nevertheless, they project that pulse harvests of the world's residual forest stock will provide the majority of industrial timber for the foreseeable future. There is no similar aggregate analysis for worldwide consumption by subsistence communities. In two self-contained local markets where deforestation has been extreme, the situation appears to have moderated. Bluffstone (1995) found that deforestation is stable in Nepal, and Hyde and Seve (1993) projected that plantation investments may offset forest removals in Malawi within ten years. Therefore, continued worldwide deforestation is not as critical an issue as it sometimes seems, either for market-based industrial timber values or for subsistence household values for a wide variety of forest resources.

Spillover Policy Effects

Policies designed for other sectors of the economy often spill over to increase deforestation. Macroeconomic policies that discourage either long-term investments and or a preference for holding real assets, policies that encourage agricultural production, and policies that affect the establishment and transfer of land tenure are common examples.

Forestry's low value means that policies with small economic impacts on other sectors tend to have relatively large unintended spillover effects on forest land use. In terms of figure 1, the shallow slopes of the land-use value functions mean that small increases or decreases in the levels of any of these functions cause larger shifts to the left or right in the location of points A and B and therefore larger shifts in forest land use. Thus, the effects of policy spillovers may be more notable in terms of hectares of forest than in terms of net forest value.

A comparison of Chile and Argentina provides a good example of macroeconomic policy impacts. Large parts of Chile and Argentina have similar forest-growing conditions. Argentina has the advantage of proximity to their mutual European markets. (Chilean exports must pay the duty for passage through the Panama Canal.) Nevertheless, Chile is a major wood exporter, while Argentina's wood exports are almost nonexistent. Argentina's high inflation and real interest rates and its unstable macroeconomic policies discourage long-term investments in all sectors, including forestry. Chile's stable macroeconomic environment permits long-term planning with confidence, and one result is extensive investment in plantation forests. Chile's macroeconomic policies raise the value of forest plantations, provide an earlier incentive for forest plantations, and expand the region of successful plantations, marked in figure 1 by the neighborhood around point B.[3]

For many countries, subsidies on agricultural inputs and price supports on agricultural outputs ensure that the market value of agricultural land is greater than its social optimum. This shifts point A in figure 1 to the right, increasing the land area devoted to agriculture, extending the amount of time before forest values are competitive with agriculture, delaying investments in trees, and decreasing the number of hectares of tree crops. There are few empirical analyses of the effects of agricultural policies on the forestry sector, but extensive worldwide intervention in agriculture supports an argument that agricultural policies can be important sources of deforestation.

Rules for establishing land tenure are the third example of policies designed for other purposes that can have an unusual effect on forestry. Any government's formal rules of tenure are designed to be general to any land use or claimant. The costs they impose are relatively constant per hectare for any land use; thus their relative impacts are often greatest on low-value land uses such as forestry. There is no problem where a country is still developing, the rules are informal, and local customs for recognizing land claims arise on their own. This has been the experience in the agricultural conversion of forest land in West Africa, and the resulting land use patterns ar efficient (Migot-Adhola and others 1991).

The legal codes of industrial countries and developing countries in Asian wit long histories of formal government structure spell out the rules for establishing tenure. These codes often constrain the initial allocation of claims on the fores frontier, but the final allocation is efficient wherever the code permits initi claimants to transfer the land to higher-value uses. Libecap and Johnson (197 showed that this was the experience for western U.S. expansion in the nine teenth century. Government policy required the initial claimants to be sma farmers, but these claimants had full rights to exchange their new property aft some minimal period, and many farmers (and speculators who called them selves farmers) quickly sold their rights to other economic enterprises.

Often, however, the formal rules of tenure establish preference among init claimants and restrict land transfers by the claimants. For example, many cou tries permit claims on the forest frontier only with evidence of capital improv ments such as fences and forest removal. This rule prevents claims for natu forest management. Binswanger (1989), Mahar (1989), and Schneider (199 for example, showed that preferential treatment for the livestock industry (a the implicit restriction on comparable claims for forest management) played important role in Amazonia's deforestation in the 1980s.

Restrictions on land transfers can be an important problem in countries w established formal institutions but without the means to enforce the formal clai on the forest. For example, Feder and others (1988) showed that the inability Thailand's Royal Forestry Department to enforce government claims on t forest, together with its reluctance to transfer land to trespassing (and probab higher-valued) users, created inefficiencies in land use. In general, governme restrictions on transfers to high-valued agricultural uses also remove the inc

tives for conservation practices and long-term agricultural management by settlers at the forest fringe in many Asian countries. Because the forest ministries are ineffective in restricting illegal access to the forest, squatters and other trespassers engage in short-term agricultural and forestry practices that increase deforestation and increase erosion (Cruz, Francisco, and Conway 1988; Amacher and others 1995).

Two hypotheses emerge from this discussion. First, restoring macroeconomic stability and correcting policy failures that inadvertently spill over to affect forestry may have a greater positive effect on the forestry sector and on the amount of residual land that remains in forest than do all preferential forestry sector policies. Second, permitting transfers of land use rights to the highest-valued use and permitting future land exchanges as values eventually instruct would improve long-term land use, limit erosion, and improve the conditions of indigenous peoples. Neither hypothesis suggests a total solution; both suggest that current policy distortions are worse than an unfettered market.

Forest Rent

The magnitude and allocation of forest charges (rent) are subjects of some contention. Gillis (1988) and Vincent (1990) argue that forest rent is large, that it belongs to the government, and that its assignment makes a difference in efficient levels of output. Paris and Ruzicka (1991) and Hyde and Sedjo (1992) argue that its magnitude varies from case to case and that its best allocation (to the government or to forest concessionaries) is a distributive matter of undetermined merit, but one that does not alter efficient land use. These arguments are valid for the publicly owned share of the forest. The rent in question is generally associated with mature natural forests and commercial harvests (in the neighborhood of point B in figure 1). Net resource values in this region are too low to justify long-term management. The fact that most public forest lands went unclaimed by earlier commercial interests is evidence that the rents accruing to public forest management are not always large and economically important.

Where the forest ministry claims the property rights and charges royalties for resource extraction, it is often the case that either the minimum fee is too high and there are no bidders or that the ministry must subsidize harvests. Both conditions suggest a low value for standing forest resources and point to the financial efficiency of open-access management. The U.S. Forest Service, for example, recently noted that 62 of its 156 national forests suffer net financial losses on overall timber sales. Many more national forests suffer net losses on individual timber sales. In other examples, the Philippine Bureau of Forestry Development operates at an overall net loss on its timber sales (Boado 1988), and the industrial forests of northeastern China suffer a similar shortfall. Maintaining mill employment is the key objective in northeastern China, and the government

subsidizes industrial forests to ensure harvest levels sufficient to protect mill employment.

If this argument is correct, it also begs the question of why the payments for some new timber concessions are so high. Where large rents are apparent for lands at the margin of previous forest resource extraction, they must be created from some recent relaxation of a restriction on forestry activity. In this case, the rent is truly a return to relaxing the restriction. (Deacon 1994 and 1995 more broadly suggests that these rents, and deforestation in general, are largely determined by the uncertainty of widely variable political conditions.) If the forest had value without the restriction, it would have been harvested earlier when its net value was newly positive and still small. Surely all private operators would have had sufficient incentive to harvest each year all the way to the geographic point where their access and harvest costs depleted the entire value of the standing resource. Rents in subsequent years can appear only where previous extraction was incomplete, and incomplete extraction would occur only where harvest policies were restrictive.

What kinds of restrictions create rents? Limited access to the forest and public policies that make harvest uneconomical are two examples. As for the former, much of the forest in northeastern Thailand, Liberia, the Amazon, and the interior of Canada and the United States would not be open for harvest had roads not been built, and forest resource values alone often cannot justify these roads. The "allowable cut" policy that restricts timber harvests to biological timber rotations is an example of the latter. The allowable cut policy is taught at Dehra Dun, India, and Los Banos, Philippines, in two of the oldest and largest forestry schools in the world. It is common to western European and North American forest ministries and probably to most of the forest ministry officials and schools around the world.

This perspective should change the debate on forest rent. In sum, forest rent is seldom large, and it is often smaller than expected because analysts overlook the costs of administering timber sales. In those important but specialized cases where the rent is large, it begs questions about the forest management activity or policy change that created the rent. Does the same policy exist elsewhere? What are its advantages? Could more rents be captured by applying the same management activity or the same change in forest policy to other timbersheds and other forests in the same region? If so, should that activity or policy be expanded?

Forestry Research and the Effect of New Technologies

The theme of forestry as a low-value and low-cost resource continues as an important explanatory factor, this time for its effect on locating successful opportunities for research and technical change. Successful research is defined by technological breakthroughs and the adoption of new technologies that lower

production costs. Yet production costs are already very low throughout most forestlands. Therefore, the opportunities for research and technological change are small. Indeed, too often, the experimental successes of forestry research are uneconomic in broad field applications because the costs of new, higher biological yields are greater than the costs of harvesting natural forest resources. Consider the potential impacts of research and the likely adoption of new technologies for the five categories of forest land use.

- *Residual forest land.* Shifting cultivation and parks for recreational uses are the only productive activities. Only research on subsistence agriculture or research that extends nonconsumptive recreational and aesthetic opportunities from a given preserved area can affect them; research on production forestry cannot.
- *Degraded open access forests and mature natural forests.* Production costs are zero in these regions. The only costs are those relating to access and to the harvest operation itself. Therefore research impacts in this region are restricted to technologies such as improved saws or cheaper transportation systems that reduce harvest or access costs, or to institutional changes that decrease the costs of establishing and protecting secure rights to the land. The technologies serve to extend the region of current harvests farther into the residual forest (or farther to the right in figure 1). The institutional changes enable land conversion from the region of open-access pulse harvests to sustainable forest plantations.
- *Private tree plantings and commercial plantations.* Only in these regions can growers take advantage of the full range of new cost-saving forest production technologies. In addition, wood saving technologies such as improved stoves (in subsistence economies) and improved wood utilization practices in the mill (for industrial forestry) are candidates for adoption in regions where the price of wood is high enough to justify investments in plantations. (See Scherr 1995 and Patel, Pinckney, and Jaeger 1995 for findings on household plantations in Kenya; Amacher, Hyde, and Rafiq 1993 on seedling distribution in Pakistan; Amacher, Hyde, and Joshee 1992 on stoves in Nepal; and Barnes and others 1993 on the adoption of improved stoves in general.)

The policy instructions indicate that biological research must focus on the species and forest types that are appropriate for industrial and smallholder plantings—and only in locations where such planting will soon or already occur. Fortunately, the increasing practice of farm forestry indicates that the opportunities for biological forestry research are increasing. Biological research should refrain from activities designed to improve species and forest products characteristic of the areas of mature and open-access natural forests. New biological technologies for these regions will not be adopted.

Social science research and institutional change, of course, may still offer possibilities in the regions of open-access harvests that might become pri-

vate property or well-managed commons. The research objective should be to find cheaper ways to establish and protect the full range of long-term property rights to these forests, including the right to legal transfer. Successful applications of this research would probably be concentrated in countries where established and inflexible formal institutions hamper private land claims and transfers, or in any country with serious erosion problems or important sources of biodiversity.

Conclusions and Final Observations

Market responses to commercial forest values and subsistence household responses to available forest resources create limits to potential deforestation. Both economic intuition and empirical evidence from a broad array of industrial and developing countries, temperate and tropical, support this argument. Furthermore, policy interventions to correct problems associated with forest land tenure, deforestation, and forest management do not necessarily improve on market-based solutions because forest resources and forest land generally have low values and are widely dispersed. Finally, forest policy interventions are often ineffective because they require extensive monitoring and administrative costs—which the low-value resources at risk cannot support.

This is a modest argument, and the basic points should be simple and clear. They are important only because today's policy environment overlooks them all too often. This encourages public forest management and broad based forest policy interventions, despite general recognition of the historic inefficiencies of forest ministries and previous forestry regulations. Perhaps it is time to recognize that some of the merits of deregulation and structural adjustment commonly acknowledged by policymakers in other sectors also pertain to forestry.

How then do we identify the targets for further inquiry and useful intervention in forest policy? Such interventions typically occur at the margins of the various land-use categories. The first target is the collection of existing macroeconomic and agricultural policies that spill over to discourage forest investment. The second is the assortment of existing policies for establishing secure rights to forest land and for transferring these rights. Many formal rights to forest resources are static and final. They are unresponsive to economic changes and to the developing economic validity of private (or well managed common) ownership of natural forests. These policies also delay socially optimal forest investments and conservation practices. Finally, both erosion of forested agricultural lands and protection of genetic reserves may require more aggressive policy proscriptions. They call for interventions that constrain market solutions, but only in specific locations where erosion extends beyond the boundaries of a single ownership or where preserving a natural forest ecosystem is desirable. It is a myth that all forests everywhere support multiple and significant nonmarket values.

Self-correcting adjustments to scarcity may create two new problems, particularly for the poorest households in the lowest-income countries. Once local subsistence farmers begin planting trees, they must forgo other production on their scarce household land. The poorest landowners may be forced into trading trees for nutrition when they give up agricultural production for trees. Landless households may suffer even more because they do not have the option of planting trees on their own lands. Expanding markets for contract labor may solve the latter problem (Bluffstone 1995), but where the markets for contract labor are not expanding, the only options for landless households may be to trespass to obtain forest products or to incur the opportunity costs of collecting from the ever-more-distant open-access forests. The alternative scenario is that some landless households will supply forest labor, and the value of their labor will increase as the value of deforestation and forest products increase. In either scenario, the status of the landless, and not deforestation, may be the most serious result of increasing forest scarcity—and it is relatively unexamined as yet.[6]

Notes

William F. Hyde is professor and Gregory S. Amacher is assistant professor in the Forestry Department at Virginia Polytechnic Institute and State University. William Magrath is senior economist in the Agriculture and Forestry Production Systems Division of the World Bank. Marti Pillo, Marian S. de los Angeles, Luis Constantino, and Robert Anderson shared critical insights during the preparation of this article. Earlier versions were presented at the annual meeting of the International Association for the Study of Common Property in Manila, June 15–19, 1993, and the European Association of Environmental and Resource Economists in Dublin, June 22–24, 1994.

1. Several recent assessments confirm this model for a relationship between access and forest cover in Belize (Chomitz and Gray 1994), Bolivia (Robbins, Kenney, and Hyde 1995), the Philippines ([illegible], Iverson, and Brown 1993), and Tanzania (Hofstad 1995).

2. Nevertheless von Thunen (1875) also places forest plantations closer than agriculture to the town and the local value center. For isolated German states in 1875, as for some subsistence agricultural communities in the 1990s, land used for these plantations had a higher net value for fuelwood production than for agriculture.

3. Harvests from the public forests are a more contentious issue. The general view is that harvests have been too slow in some regions, too rapid in others, and seldom at rates justified by either private or social criteria. See Berck (1979), Hyde (1980), and Repetto (1988).

4. The analyses of general macroeconomic and trade policy effects on the forestry sector are scant. Binswanger (1989) identified a variety of fiscal and monetary policies that encouraged deforestation in the Brazilian Amazon. Browder (1987) observed that Brazil's general export promotion policy encouraged a predatory system of advance timber purchasing that, for one company alone, extended harvests on marginal forest land in Rondonia by 300,000 hectares. Boyd and Hyde (1989) concluded that in the United States a favorable capital gains tax policy had more effect than direct forest sector policies.

5. "Allowable cut" has many meanings. It variously can mean long- or short-run planned timber sales, successful bids for a harvest over a period of time, the model for determining a "planned" sale, or reported harvests for a preceding year. The term must be used with care. The general model taught in forestry schools and applied by forest ministries around the world begins with a biological definition of productive timberland that generally overstates the economic forest land base. It estimates harvests from these lands as some variant

of the volume or area of mature forest divided by the harvest age at biological maturity, plus average annual growth for the full area. Biological rotations depend on the species, but they generally exceed economic rotations by 35 percent or more. For land that is economic, this model sharply constrains harvests. Because all land in the analysis is not economic, the net impact on harvest levels is uncertain and varies from case to case. The net economic effect is always negative. See Davis and Johnson (1987) or Hyde (1980) for statements of the allowable cut model and Hirshleifer (1974), Samuelson (1976), and Hyde (1980) for detailed critiques.

6. The exceptions in the economic literature are Kumer and Hotchkiss (1988), Barnes and Qian (1994), Larson and others (1996), and the work of N. S. Jodha, summarized in Jodha (forthcoming).

References

The word "processed" describes informally reproduced works that may not be commonl available through library systems.

Aarne, Marti, ed. 1995 *Yearbook of Forest Statistics* Helsinki Finnish Forest Researc Institute.

Amacher, Gregory S., Luis Constantino, William F. Hyde, and Monica Rios 1996 "Mac roeconomic Effects of Natural Resource Enhancement in Small Open Economies Chile Forest Incentive Program " Department of Forestry, Virginia Polytechnic Institute an State University, Blacksburg, Va Processed

Amacher, Gregory S., M. Conception Cruz, Wilfredo Cruz, and William F. Hyde 199 "Environmental Motivations for Migration Population Pressure, Poverty, and Defor estation in the Philippines " Department of Forestry, Virginia Polytechnic Institute a State University, Blacksburg, Va Processed

Amacher, Gregory S., William F. Hyde, and Bharat R Joshee 1992 "The Adoption Consumption Technologies under Uncertainty The Case of Improved Stoves in Nepa *Journal of Economic Development* 17(2):93-105

———. 1993. "Joint Production and Consumption in Traditional Households Fuelwoo and Agricultural Residues in Two Districts of Nepal " *Journal of Development Studi* 30(1):206-25.

Amacher, Gregory S., William F. Hyde, and Keshav R Kanel Forthcoming "Househ Fuelwood Demand and Supply in Nepal Choice between Cash Outlays and Lab Opportunity " *World Development*

Amacher, Gregory S, William F Hyde, and Mohammed Rafiq 1993 "Local Adoption New Forestry Technologies With an Example from Pakistan's Northwest Frontier Pr ince." *World Development* 21(3):445-54

Anderson, Dennis 1987 *The Economics of Deforestation* Baltimore Johns Hopkins U versity Press.

Bailey, Ronald, ed. 1995 *The True State of the Planet* New York Free Press

Barnes, Douglas F., Keith Openshaw, Kirk R Smith, and Robert van der Plas 1993 "I Design and Diffusion of Improved Cooking Stoves " *World Bank Research Obse* 8(2):119-41.

Barnes, Douglas F., and Liu Qian. 1994. "Social Impacts of Energy Policies " In Doug Barnes, J. Dowd, Liu Qian, K Krutilla, and William Hyde, eds , "Urban Energy Tr sitions, Poverty, and the Environment." World Bank, Industry and Energy Departme Washington, D C . Processed.

Barnett, Harold J., and Chandler Morse. 1963. *Scarcity and Growth. The Economics Natural Resource Availability*. Baltimore: Johns Hopkins University Press for Resourc for the Future.

Bensel, Terrence G., and Elizabeth M. Remedio. 1993. "Patterns of Commercial Woodfuel Supply, Distribution, and Use in the City and Province of Cebu." Regional Wood Energy Development Programme Working Paper. Food and Agricultural Organization, Bangkok. Processed.

Berck, Peter. 1979. "The Economics of Timber: A Renewable Resource in the Long Run." *Bell Journal of Economics* 10(2):447–62.

Binswanger, Hans. 1989. "Brazilian Policies That Encourage Deforestation in the Amazon." Working Paper 16. World Bank, Environment Department, Washington, D.C. Processed.

Bluffstone, Randall. 1995. "The Effect of Labor Markets on Deforestation in Developing Countries under Open Access: An Example from Rural Nepal." *Journal of Environmental Economics and Management* 29(1):42–63.

Boado, Eufresina L. 1988. "Incentive Policies and Forest Use in the Philippines." In Robert Repetto and Malcolm Gillis, eds., *Public Policy and the Misuse of Forest Resources.* Cambridge, U.K.: Cambridge University Press.

Boyd, Roy G., and William F. Hyde. 1989. *Forestry Sector Intervention. The Impacts of Public Regulation on Social Welfare.* Ames, Iowa: Iowa State University Press.

Browder, John O. 1987. "Brazil's Export Promotion Policy (1980–1984): Impacts on the Amazon's Industrial Wood Sector." *Journal of Developing Areas* 21:285–304.

Byron, Neil. 1984. "People's Forestry: A Novel Perspective of Forestry in Bangladesh." *ADAB News* (Association of Development Agencies in Bangladesh) 11(2):28–35.

———. 1993. Personal records.

Chomitz, Kenneth M., and David A. Gray. 1994. "Roads, Land, Markets, and Deforestation: A Spatial Model of Land Use in Belize." Policy Research Working Paper 1444. World Bank, Policy Research Department, Washington, D.C. Processed.

Clawson, Marion. 1979. "Forests in the Long Sweep of American History." *Science* 204:1168–74.

Cline-Cole, R. A., H. A. C. Main, and J. E. Nichol. 1990. "On Fuelwood Consumption, Population Dynamics and Deforestation in Africa." *World Development* 18(4):513–27.

Cruz, Wilfrido D., H. A. Francisco, and Z. T. Conway. 1988. "The Onsite and Downstream Costs of Soil Erosion in the Magat and Pantabangan Watersheds." *Journal of Philippine Development* 15(1):48–88.

Davis, Lawrence S., and K. Norman Johnson. 1987. *Forest Management.* New York: McGraw-Hill.

Deacon, Robert T. 1994. "Deforestation and the Rule of Law in a Cross Section of Countries." *Land Economics* 70(4):414–30.

———. 1995. "Assessing the Relationship between Government Policy and Deforestation." *Journal of Environmental Economics and Management* 28(1):1–18.

Dewees, Peter A. 1989. "The Woodfuel Crisis Reconsidered: Observations on the Dynamics of Abundance and Scarcity." *World Development* 17(8):1159–72.

Eckholm, Eric P. 1976. *Losing Ground: Environmental Stress and World Food Prospects.* New York: W. W. Norton.

Eggertsson, Thrainn. 1990. *Economic Behavior and Institutions.* New York: Cambridge University Press.

Feder, Gershon, Tongroj Onchan, Y. Chalamwong, and C. Hongladarom. 1988. *Land Ownership Security, Farm Productivity, and Land Policies in Thailand.* Baltimore: Johns Hopkins University Press.

Gautam, K. H., B. R. Joshee, R. L. Shrestha, V. K. Silwal, M. P. Suvedi, and L. P. Uprety. Forthcoming. "Some General Features of Household Forestry in Nepal." In William F. Hyde and Gregory S. Amacher, eds., *Social Forestry: An Empirical Introduction from South and Southeast Asia.* New Delhi: Oxford and IBH.

Gillis, Malcolm. 1988. "Indonesia: Public Policies, Resource Management and the Tropical Forest." In Robert Repetto and Malcolm Gillis, eds., *Public Policy and the Misuse of Forest Resources.* Cambridge, U.K.: Cambridge University Press.

Hirshleifer, Jack. 1974. "Sustained Yield Versus Capital Theory." In Barney Dowdle, ed., *Proceedings of a Conference on the Economics of Sustained Yield Forestry.* Seattle: University of Washington, College of Forestry.

Hofstad, Ole. 1995. "Woodland Deforestation by Charcoal Supply to Dar es Salaam." Agricultural University of Norway, Forestry Department, As. Processed.

Hyde, William F. 1980. *Timber Supply, Land Allocation, and Economic Efficiency.* Baltimore: Johns Hopkins University Press.

Hyde, William F., and James L. Chamberlain. 1995. "Who Gains from Privatizing the U.S. Forest Service?" *Journal of Forestry* 94(8):22–26.

Hyde, William F., and Roger A. Sedjo. 1992. "Managing Tropical Forests: Reflections on the Rent Distribution Discussion." *Land Economics* 68(3):343–50.

Hyde, William F., and Juan Seve. 1993. "The Economic Role of Wood Products in Tropical Deforestation: The Severe Example of Malawi." *Forest Ecology and Management* 57(2):283–300.

Jodha, N. S. Forthcoming. "Common Property Resources and the Dynamics of Rural Poverty: Field Evidence from the Dry Regions of India." In William F. Hyde and Gregory S. Amacher, eds., *Social Forestry: An Empirical Introduction from South and Southeast Asia.* New Delhi: Oxford and IBH.

Johnson, Ronald N., and Gary D. Libecap. 1980. "Efficient Markets and Great Lakes Timber: A Conservation Issue Reexamined." *Explorations of Economic History* 17:372–85.

Kanel, Keshav R. 1993. "Market Opportunities in Nepal's Tarai." Report prepared for Finnish International Development Authority, Kathmandu. Processed.

Krutilla, Kerry, William F. Hyde, and Douglas F. Barnes. 1995. "Urban Energy Consumption and Periurban Deforestation." *Forest Ecology and Management* 74(2):181–95.

Kumer, Shubh K., and David Hotchkiss. 1988. "Consequences of Deforestation for Women's Time Allocation, Agricultural Production, and Nutrition in Hill Areas of Nepal." Research Report 69. International Food Policy Research Institute, Washington, D.C. Processed.

Larson, Bruce A. 1992. "Changing the Economics of Environmental Degradation: Lessons from the National Environmental Action Plan Process." Winrock International Working Paper. Washington, D.C. Processed.

Larson, Bruce A., Glenn Rogers, Joy Green Larson, and Grenville Barnes. 1996. "Access to Forests and Child Health in Zaire." *Forest Science* 42(1):3–9.

Libecap, Gary D., and Ronald N. Johnson. 1979. "Property Rights, 19th Century Federal Timber Policy, and the Conservation Movement." *Journal of Economic History* 39(1):129–42.

Liu, Dawning S., Louis R. Iverson, and Sandra Brown. 1993. "Rates and Patterns of Deforestation in the Philippines." *Forest Ecology and Management* 57(1):1–16.

Mahar, Dennis J. 1989. "Government Policies and Deforestation in Brazil's Amazon Region." Working Paper 7. World Bank, Environment Department, Washington, D.C. Processed.

McConnell, Kenneth E. 1983. "An Economic Model of Soil Conservation." *American Journal of Agricultural Economics* 64(1):83–89.

Mendelsohn, Robert, and Matthew Balick. 1995. "The Value of Undiscovered Pharmaceuticals in Tropical Forests" *Economic Botany* 4(2):223–28.

Mercer, Evan, and John Soussan. 1992. "Fuelwood Problems and Solutions." In N. P. Sharma, ed., *Managing the World's Forests.* Dubuque, Iowa: Kendall/Hill.

Migot-Adholla, Sham, Peter Hazell, Benoit Barel, and Frank Place. 1991. "Indigenous Land Rights Systems in Sub-Saharan Africa: A Constraint on Productivity?" *World Bank Economic Review* 5(1):155–75.

Miranowski, John A. 1984. "Impact of Productivity Losses on Crop Production and Management in a Dynamic Economic Model." *American Journal of Agricultural Economics* 66(1):61–71.

Paris, Remy, and Ivan Ruzicka. 1991. "Barking Up the Wrong Tree: The Role of Rent Appropriation in Sustainable Tropical Forest Management." Environment Office Occasional Paper 1. Asian Development Bank, Manila. Processed.

Patel, H. Sandeep, Thomas C. Pinckney, and William K. Jaeger. 1995. "Smallholder Wood Production and Population Pressure in East Africa: Evidence of an Environmental Kuznets Curve?" *Land Economics* 71(4):516–30.

Repetto, Robert. 1988. "Subsidized Timber Sales from National Forest Lands in the United States." In Robert Repetto and Malcolm Gillis, eds., *Public Policy and the Misuse of Forest Resources*. Cambridge, U.K.: Cambridge University Press.

Republic of the Philippines. 1987. *Philippine Statistical Yearbook*. Manila: National Economic Development Authority.

Rice, Richard. 1993. Personal communication.

Robbins, Eugene S., Robert N. Kenny, and William F. Hyde. 1995. "Forest Concessions and Industrial Policy in Bolivia." Report to the Undersecretariat of Natural Resources/Bolfor, La Paz. Processed.

Ruttan, Vernon W., and J. C. Callahan. 1962. "Resource Inputs and Output Growth: Comparisons between Agriculture and Forestry." *Forest Science* 8(1):68–82.

Samuelson, Paul A. 1976. "Economics of Forestry in an Evolving Society." *Economic Inquiry* 14:466–92.

Scherr, Sara J. 1995. "Economic Factors in Farmer Adoption of Agroforestry: Patterns Observed in Western Kenya." *World Development* 23(5):787–804.

Schneider, Robert. 1993. "Land Abandonment, Property Rights, and Agricultural Stability in the Amazon." LATEN Dissemination Note 3. World Bank, Latin America and the Caribbean Technical Department, Environment Division, Washington, D.C. Processed.

Sedjo, Roger A. 1992. "Property Rights, Genetic Resources, and Biotechnological Change." *Journal of Law and Economics* 35:199–213.

Sedjo, Roger A., and Kenneth S. Lyon. 1990. *The Long-Term Adequacy of World Timber Supply*. Washington, D.C.: Resources for the Future.

Simpson, David, Roger A. Sedjo, and John Reid. Forthcoming. "Valuing Biodiversity for Use in Pharmaceutical Research." *Journal of Political Economy* 104(1).

Stevens, Stan. 1988. "Sacred and Profaned Himalayas." *Natural History* 97(1):26–35.

Templeton, Scott. 1992. "Major Determinants of Farmer Use of Semi-Permanent Soil Erosion Controls in Upland Areas of the Philippines." Ph.D. dissertation. University of California, Berkeley, Agricultural Economics Department. Processed.

U.S. Department of Agriculture. 1992. *Agricultural Statistics*. Washington, D.C.: Government Printing Office.

Vehkamaki, Seppo, and Markku Simula. 1995. "Economic Effects of Biodiversity in Forests." Paper presented at the International Union of Forest Research Organizations World Congress. Helsinki University and Finnish Forest Research Institute, Tampere, Finland. August. Processed.

Viana, Virgilio M., Ricardo A. Mello, Luiz M. deMorais, and Nilson T. Mendes. 1996 Ecology and Management of Brazil Nut Populations in Extractive Reserves in Xapuri, Acre." Departamento Ciencias Florestais, Universidade de São Paulo, Brazil. Processed.

Vincent, Jeffrey. 1990. "Rent Capture and the Feasibility of Tropical Forest Management." *Land Economics* 66(2):212–23.

Vincent, Jeffrey R., and Clark S. Binkley. 1992. "Forest-Based Industrialization: A Dynamic Perspective." In N. P. Sharma, ed., *Managing the World's Forests.* Dubuque, Iowa: Kendall/Hill.

von Thunen, Johann H. 1875. *The Isolated State in Relation to Land Use and National Economy* Berlin: Schmaucher Zarchlin.

World Bank. 1992. *World Development Report 1992: Development and the Environment.* New York: Oxford University Press.

World Resources Institute. 1994. *World Resources, 1994–95: A Guide to the Global Environment.* New York: Oxford University Press.

Yin, Runsheng. 1994. "The Farmer Tree Planting Response to Political Uncertainty in China." Ph.D. dissertation. University of Georgia, College of Forest Resources, Athens Processed.

Yin, Runsheng, and William F. Hyde. 1995. "Trees as an Agriculture Sustaining Resource: Evidence from Northern China." University of Georgia, College of Forest Resources, Athens. Processed.

FINANCIAL MARKETS, PUBLIC POLICY, AND THE EAST ASIAN MIRACLE

Joseph E. Stiglitz
Marilou Uy

Many factors contributed to the rapid growth of the economies of East Asia in the past quarter century. This article examines one important aspect of that growth—commonly referred to as the "East Asian miracle"—public policies affecting the financial markets. East Asian governments intervened extensively in financial markets at all stages of their development. What sets their actions apart from those of other developing countries that have not fared as well? We do not have the information to answer conclusively what effect particular actions had (that requires a counterfactual test of what growth would have been without the particular intervention). But we can identify the market failures the East Asian governments addressed, assess some of the theoretical reasons why each policy might be growth enhancing, and provide some data attesting to the impacts of the policy. Several characteristics of financial sector interventions in East Asia stand out. they incorporated design features that improved the chances of success and reduced opportunities for abuse, interventions that did not work out were dropped unhesitatingly; and policies were adapted to reflect changing economic conditions.

The economies of East Asia, beginning with Japan and followed by the Four Tigers—the Republic of Korea, Hong Kong, Singapore, and Taiwan (China)—and then by Malaysia, Indonesia, and Thailand, have grown so rapidly during the past quarter century that their growth has been called the "East Asian miracle." What lessons for other countries can be derived from this experience, assuming that it was not really a miracle, but rather a consequence of some well designed programs and policies? In particular, what can be learned from the policies affecting financial markets? East Asian governments have intervened intensively in the operations of their financial systems. They have helped create financial markets and institutions, regulated them heavily, and directed credit to some indus-

The World Bank Research Observer, vol. 11, no. 2 (August 1996), pp. 249–76

tries and away from others. These actions have been intended to mobilize savings and to affect the allocation of investment, but their effects have extended well beyond the capital market: the "prize" of scarce credit, awarded for good export performance, has provided strong investment incentives.

Careful examination of the interventions provides insights into the importance of specific design features that increased the likelihood that the interventions would work and decreased the likelihood of abuses.[1] Government flexibility was also important. When programs failed, the interventions behind them were abandoned, and as economies changed, so did the role of governments. These general precepts, more than the particular interventions adopted, may prove to be the most important insights that other developing countries can borrow from the East Asian miracle.

Government interventions were directed at two broad objectives. The first is straightforward: making financial markets and institutions work better. Without the intermediation provided by capital markets, firms would have to rely solely on retained earnings for their investments, and firms' expected marginal private returns from investment would diverge markedly, especially in the short run. When capital markets work well, marginal returns are equated in all sectors and firms. Moreover, by spreading and pooling risks more broadly, capital markets lower risk premiums, so firms can undertake investments with greater risks and higher expected returns.

But even if the marginal private returns from investment are equated in all sectors and firms, capital may not be allocated efficiently if there are systematic deviations between private and social returns. Thus the second objective of government intervention in financial markets is to correct any resulting misallocation of resources.

Government financial policies are of three types: creating markets and financial institutions; regulating them; and providing rewards (subsidies or access to credit or foreign exchange, often on preferential terms) to firms, groups, or industries that undertake priority activities or perform in an exemplary manner. The East Asian economies had high national saving rates, achieved largely by voluntary actions, and they were able to invest their savings in ways that yielded high returns. Government interventions in the financial market that promoted savings and the efficient allocation of capital were central to these successes. Five of the more important interventions are examined here: promoting savings, regulating banks to fortify their solvency, creating financial institutions and markets, enforcing financial restraint, and intervening directly in the allocation of credit.

Promoting Savings

East Asian governments promoted national saving in several ways, from creating financial institutions and regulating them to running small fiscal deficits or even surpluses.

Creating Postal Savings and Provident Funds

The postal saving systems in Japan, Malaysia, Singapore, and Taiwan (China) were the most important of the institutions governments created to promote savings. These systems attracted multitudes of small savers by giving them security and convenient access. When Japan's postal saving system was created in 1875, other financial institutions had commonly excluded small savers or discouraged them by paying low interest rates or requiring minimum deposits too high for small savers (Mukai 1963). The postal saving system also provided convenient access through an extensive branch network of post offices, especially in rural areas (Yoshino 1992, Shea 1993, and Chiu 1992). Japan further encouraged postal savings by exempting the interest paid on postal deposits from income tax (until 1988); Taiwan (China) since 1965 and Korea do the same below a certain threshold.

The postal saving banks mobilized huge amounts of saving—up to 25 percent of national saving in Japan since the 1950s, 20 percent in Taiwan (China), and 12 percent in Singapore. In Japan postal saving was particularly important during periods of financial distress, when confidence in commercial banks waned. During the 1920s, for example, when the banks became unstable, households shifted their savings from banks to postal savings. In 1920 demand deposits in banks were six times the amount held in postal savings; by 1930 they had dwindled to just twice that amount.

Although the positive role of postal saving is generally acknowledged, the compulsory pension plans of Malaysia and Singapore are more controversial. Two rationales are given for such plans. First, most developing countries lack private annuity markets, and the few that exist are barely functional. Indexed annuities, for example, are rarely available, and nonindexed annuities are often priced unattractively because of large transactions costs and problems of adverse selection. In Malaysia and Singapore not only did the government provide the annuities, but it also required individuals to participate in the plans. This leads to the second rationale: even when annuities are available, people may not save enough for their old age, and governments end up assisting elderly people who are unable to support themselves. To avoid this free-rider problem, government may require citizens to have at least a minimum level of savings for retirement.

Unlike the social security systems of Europe and the United States, East Asia's programs were fully funded rather than pay-as-you-go. The distinction is important because fully funded systems are more likely to increase the national saving rate. Also setting them apart from U.S. and European systems were the extremely high required contribution rates (up to 50 percent of salaries in Singapore and 28 percent in Malaysia) and the fact that the savings covered not just retirement but other purposes as well, such as providing the down payment for housing.

The impact of provident funds on aggregate saving depends on how much voluntary saving they displace. In East Asia displacement does not appear to have been a

problem: private pension schemes grew alongside public ones as individuals pursued ways to support increased consumption during retirement. At worst the pension funds' negative effects were so small that aggregate saving still increased. Dekle (1990) finds little empirical support of a negative effect of pension funds on private saving in Japan, while Noguchi (1985) finds a small negative effect. Singapore's Central Provident Fund had a positive effect on aggregate saving: had provident fund contributions remained at 10 percent, their required level in 1966, instead of rising to 35 percent during the 1970s and to 50 percent during the 1980s, average saving rates over the past two decades would have been about 4 percentage points lower (Monetary Authority of Singapore 1991).

Regulating to Encourage Saving

Governments used three types of regulations to influence national saving rates. Some regulations were designed to discourage consumption, some to enhance the safety and soundness of private financial intermediaries, and some to transfer resources from households to corporations.

RESTRICTIONS ON CONSUMER CREDIT. Most East Asian governments discouraged consumption by deliberately preventing mortgage markets and other instruments of consumer credit from developing. With little or no consumer credit available to purchase housing, consumer durables, and other goods, households were forced to save the full amount if they wanted to buy a house or make other large purchases. Although demand for consumer durables increases with household incomes, consumers could not borrow to buy these goods, so savings as a share of income rose rapidly. There is evidence that constraints on liquidity and on the development of consumer credit markets have significant effects on saving and may explain some of the differences in saving rates across countries (Jappelli and Pagano 1994). Once households acquire the necessary consumer durables, their saving rates stabilize and even drop slightly.

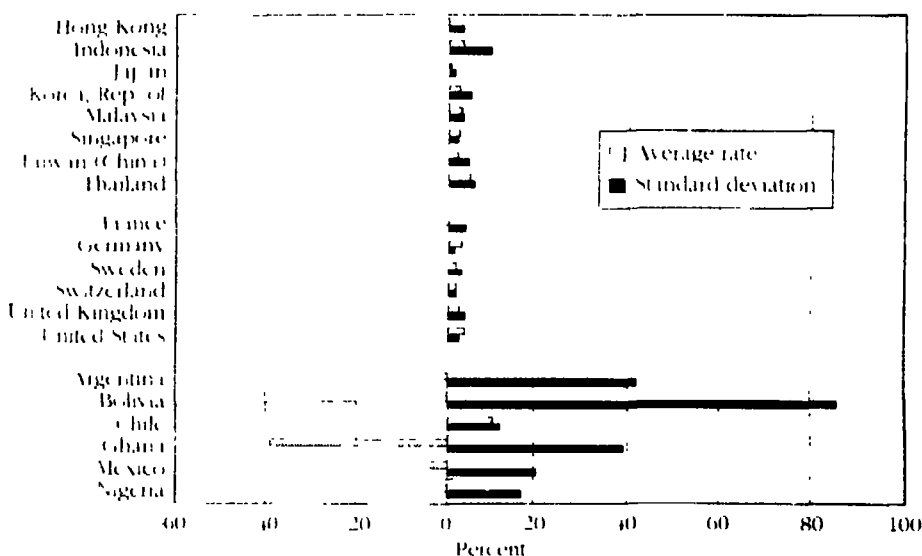
PRUDENTIAL AND OTHER BANK REGULATIONS. Prudential and other bank regulations significantly lowered the likelihood of bank failure. This increased security led to larger deposits.

FINANCIAL RESTRAINT. The conventional wisdom is that financial repression depresses saving, particularly deposits in financial institutions. But moderate repression—what this article calls *financial restraint*—may actually increase saving. Lowering interest rates transfers incomes from households to corporations and because the corporate sector has a higher propensity to save, aggregate saving increases. This positive redistribution effect is partly offset by the negative interest rate effect: as long as saving responds to changes in interest rates (the interest elasticity of saving is positive), lowering interest rates reduces household saving. But the empirical evidence suggests that this effect is small (see

Ishikawa 1987 on Japan, Nam 1991 on Korea, and Sun and Liang 1982 on Taiwan, China) and that it is overwhelmed by the redistribution effect (Balassa 1989; Giovannini 1983; Gupta 1984).

Although East Asian economies kept interest rates below equilibrium levels, financial repression was moderate compared with that practiced in many other developing countries with interest rate controls. Average interest rates in East Asia have been moderately negative at worst; many economies in other regions have had highly negative rates (figure 1).

The policies associated with financial restraint also increased saving because governments such as Korea's pitted firms against each other to see which could achieve the highest rate of exports and investment. Winners were rewarded with access to cheap credit and the rents associated with other artificially created scarcities. Achieving high rates of investment required high rates of corporate saving. Financial restraint also helped promote household saving by making the financial system more stable. Financial repression (accompanied by entry restrictions) enhanced the profitability—and thus the stability—of financial insti-

Figure 1. *Average Real Deposit Rates in Selected Countries, 1978–91*

Note: Data for Hong Kong are for 1973–91. Data for Thailand are for 1978–90. Data for Argentina exclude the extreme inflation years of 1984, 1989, and 1990. Data for Bolivia exclude the extreme inflation years of 1985 and 1986. Deposit rates for Ghana are unavailable for 1989 and 1990 and unavailable for Switzerland for 1978–80.

Source: IMF (1995).

tutions and, by boosting the franchise value of banks, provided strong incentives for banks to undertake prudent investments.

Keeping the Economy Stable and Deficits Low

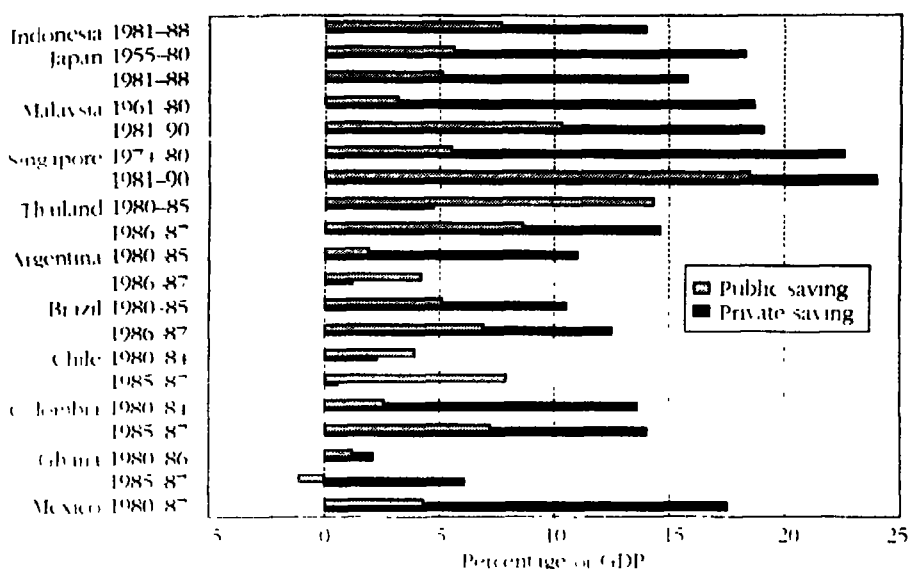
Macroeconomic policies in East Asia are far more stable than those in most other developing countries. Macroeconomic stability has a positive effect on saving for a variety of reasons. Because most countries do not have fully indexed accounts, macroeconomic stability, particularly low rates of inflation, reduces the variability of return on saving; a more secure return may increase saving. High and extremely volatile inflation, which often generates large negative real interest rates, is particularly likely to discourage saving. The East Asian economies have generally maintained positive and stable real interest rates on deposits, paralleling interest rate trends in the United States and other industrial countries (see figure 1).

Governments' tendency to run small deficits or large surpluses contributed to macroeconomic stability, and the budgetary surpluses contributed directly to high levels of national saving. All East Asian countries have maintained consistently high public saving as well as growing private saving (figure 2).

Did Interventions Increase Saving Rates?

That East Asian governments undertook a variety of actions to increase national saving is clear; what is less clear is whether saving rates were higher than they otherwise would have been. Certainly, the region's remarkably high saving rates suggest that government policies had an effect. Studies examining demographic patterns indicate that only part of this difference can be explained using standard life-cycle models. Several authors have emphasized cultural factors in explaining cross-country variations in saving (Horioka 1990). Although these factors may play a role, they do not explain why saving rates were so much higher during the past three decades than in, say, the period before World War II. Surely, entire cultures did not suddenly change.

Rapid growth is one of the distinguishing characteristics of the high-performing East Asian economies. Is it possible that high growth caused high saving, rather than the other way around? Studies have shown a strong correlation between saving and growth (Carroll, Weil, and Summers 1993) and between saving and investment and investment and growth (World Bank 1989). But these correlations do not reveal the direction of causation (nor do they rule out the possibility that other factors are the cause of the observed changes). We performed standard tests (called Granger causality tests) to ascertain whether income predicts saving or whether saving more accurately predicts income. The results show income growth to be a better predictor of saving in Indonesia, Japan, Korea, Taiwan (China), and Thailand, but the results were ambiguous for Hong Kong and Malaysia. (In Singapore income growth was not a predictor

Figure 2. *Public and Private Savings in Selected Developing Countries*

of the rapid increase in saving rates, a result consistent with a 1991 study by the Monetary Authority of Singapore showing that demographic factors and the policies of the Central Provident Fund accounted for the increase.)

Finding empirical support for a progression from income growth to saving growth is important because it suggests a virtuous cycle in which high growth leads to high saving and high saving leads to high growth. Thus government policies may be able to transform an economy with low income growth and saving to one with high income growth and saving. Moreover, government policies that enhance productive investments and raise income may have a compound effect on growth by increasing saving rates.

Advocates of less government interference in the market claim that such interventions are both ineffective and undesirable. Why interfere with individuals' preferences? There are two broad answers. First, many government interventions represent attempts to remedy market failures, to make markets work better or to create institutions to fill in gaps left by markets. In particular, the social returns to saving and investment may exceed the private returns, as is the case when there is learning from new investments and the benefits of the learning are not fully appropriable by the investor. Second, saving rates affect the

intergenerational distribution of income. Markets may yield (Pareto) efficient resource allocations under ideal circumstances, but the distribution of income—including the intergenerational distribution of welfare—yielded by markets is in no way ideal.

Regulating Banks to Enhance Their Solvency

East Asian governments have imposed several regulations to enhance the solvency of financial institutions, thereby improving both the saving rate and the efficiency of resource allocations. As noted earlier, security is as important to saving as is the rate of interest. Individuals may save even if they have no confidence in any financial institution, but the returns are likely to be lower than could be obtained through a well-functioning financial system. Moreover, well-functioning financial institutions are essential to efficient financial intermediation, channeling funds to their most efficient use in an economy.

There is ample evidence that financial crises occur with remarkable frequency in the absence of government intervention. Private monitoring apparently does not suffice to prevent a financial crisis. Moreover, no single financial institution will exercise sufficient care on its own to avoid financial distress. Information failures (individuals may not be able to sort out good banks from bad, for lack of complete information) and credit links among banks mean that the effects of financial failures may spread far beyond an individual bank in distress—there are large negative externalities. These difficulties are exacerbated by the moral hazard problems that arise with undercapitalized financial institutions, because such institutions are more likely to take large risks—they have less to lose if a loan goes bad than better-capitalized institutions. (The market failure rationale for government intervention is treated at length in Stiglitz 1994.)

Prudential Regulations to Keep Financial Institutions Sound

Japan, Hong Kong, and Singapore began strengthening prudential regulations (sometimes referred to as regulations for safety and soundness) during the 1970s; Malaysia, Taiwan (China), and Thailand followed suit in the 1980s and Indonesia in the 1990s. The stringent regulations common to these economies were adopted for three sets of reasons. Some countries, such as Singapore, recognized the importance of sound prudential regulations early on, not only for their internal capital markets but also for international commerce. Singapore has greatly benefited from its stringent regulations—financial services account for about 17 percent of its gross domestic product. The financial system's size is largely attributable to the confidence of the foreign financial and business community. Even countries that do not aspire to being a regional financial center have recognized that closer links with international markets require that domestic banks meet international standards.

Other countries imposed such regulations as part of the development process. During early stages governments owned or directly controlled banks and other financial institutions. With development governments gave up direct control. Increased deregulation (of interest rates and entry, for example) reduced governments' leverage on bank behavior, requiring stronger indirect prudential regulation in its place.

Some economies introduced prudential regulations only after experiencing financial difficulties. For example, Hong Kong strengthened its prudential regulations after the financial crises (brought on by real estate speculation) of 1965 and 1985.

Prudential regulation takes a variety of forms, each requiring a different degree of supervision by the regulator. Capital, net worth, and collateral requirements are easiest to monitor, although even they involve some degree of discretion (such as in valuing collateral). At the other extreme is judging the riskiness of particular transactions, which requires the active involvement of bank examiners.

CAPITAL ADEQUACY REQUIREMENTS. Capital adequacy standards are probably the most important tool governments can use to ensure the solvency of financial institutions. These standards make it less likely that liabilities will exceed assets and provide incentives for banks to undertake appropriate risks. The savings and loan debacle in the United States can be partly attributed to the high risks assumed by banks with low or negative net worth. All East Asian economies have adopted the capital adequacy requirements set by the Bank for International Settlements. Most economies had already imposed comparable standards on their own. The one exception is Indonesia, which implemented the new guidelines only recently.

COLLATERAL REQUIREMENTS. Regulators in East Asia have also encouraged banks to impose sizable collateral requirements to reduce the risks arising from defaults. This practice has earned East Asia's banks the reputation of being "pawnshop" banks. Although collateral requirements will not keep banks solvent, regulators have adopted this conservative system to limit banks' ability to take risks. One unintended impact of this practice has been to tie banks to the fortunes of the real estate market because most collateral has been in the form of real estate. Thus banks have overextended their lending during periods of high asset inflation, exposing themselves to greater portfolio risks during periods of declining asset value. In Japan, for example, the dramatic fall in real estate prices in the first half of the 1990s created portfolio problems for banks that had engaged in speculative real estate lending during the 1980s, when asset prices were high.

LENDING RESTRICTIONS. One of the most important objectives of East Asian policymakers was to discourage speculative lending, which has become the main

source of financial disruptions in Hong Kong, Malaysia, Thailand, and, recently, Japan (table 1). The adverse impact of speculative lending has been aggravated by bank lending to related parties, as bank owners sought to capture the gains from their speculation. In response regulatory authorities have increasingly restricted lending for real estate and to related parties—as well as lending concentrated on a few borrowers. Restrictions on related lending have been difficult to implement, however, because disclosure rules are generally poor, and in Indonesia, Japan, and Thailand, banks and firms have interlinked ownership, and companies are closely held.

DIRECT SUPERVISION. Except in Indonesia, central banks (often working with ministries of finance) in East Asia have done a good job of supervising commercial banks' loan portfolios, resulting in a lower proportion of nonperforming loans than in many other developing countries. Central banks have also supervised bank management, restraining the entry of potentially fraudulent or incompetent lenders. Singapore takes pride in the fact that its regulatory authorities detected problems in the Bank of Credit and Commerce International problems that escaped detection by allegedly more sophisticated regulators, including those in the United States and United Kingdom— and refused to allow its entry. There are major exceptions to this generally favorable picture of bank supervision, however: the financial difficulties of banks in Hong Kong, Indonesia, Korea, and Malaysia during the 1980s were brought about in part by weak supervision. Indonesia recently introduced more rigid bank supervision when it liberalized entry for private banks and as the problem of rising nonperforming loans became apparent.

The East Asian style of regulation has, at least until recently, been based more on regulatory discretion and constant interaction between regulators and banks than on the more structured rules that characterize supervision in industrial countries. This approach enables regulators to provide banks with feedback on the riskiness of their portfolios. Contemporary supervision practices in most East Asian countries (Japan, Malaysia, and Thailand are good examples) seem to combine modern prudential rules with traditional interactive monitoring, in which supervisors elicit cooperation by using the government's leverage over branch licensing, rediscounts, and other regulations. Although allowing such discretion creates the potential for abuse, abuses do not seem to be prevalent. In other countries, however, this style of regulation could create problems.

PRUDENTIAL BEHAVIOR OF GOVERNMENT BANKS. Unlike government-owned banks in many other developing economies, those in Korea (until their privatization in 1983), Singapore, and Taiwan (China) seem to have behaved prudently. In other countries political considerations often distort the lending decisions of government-owned banks. Many such banks have ended up lending to make up the losses of inefficient public enterprises. And because governments have deep pockets and tend to recapitalize public banks when they run int

financial trouble, the regulators and managers of publicly owned banks are often less concerned with solvency than private banks are. East Asian governments took several steps to minimize these problems. Taiwan (China) avoided these risks by imposing strict collateral requirements and giving the employees of public banks incentives to act prudently, going so far as to penalize employees whose loans did not perform. Korea imposed strict performance criteria to guide banks' lending decisions. In Malaysia public officials are prohibited from serving on the boards of public banks. But publicly owned banks in Indonesia and a few in Malaysia have not been properly monitored and consequently have experienced high levels of nonperforming loans.

REGULATING NONBANK FINANCIAL INSTITUTIONS. Regulators in East Asia and other developing countries have also been concerned with the appropriate regulation of nonbank financial institutions such as merchant banks, leasing companies, and cooperatives. Since the early 1980s these institutions have grown in number, but they have been less closely regulated than commercial banks, leading to several major insolvencies (see table 1). For instance, during the first half of the 1980s, Malaysia's deposit-taking cooperatives and Thailand's numerous finance companies became insolvent. These failures threatened the solvency of commercial banks, which often used nonbank financial subsidiaries as a way to avoid close scrutiny by regulators. In response to these insolvencies regulators have increased their supervision of nonbank financial institutions.

Protecting Banks from Competition

Nearly all governments regulate the entry and operations of financial institutions to ensure that new entrants and incumbents are safe and solvent. But most East Asian governments (with the exceptions of Hong Kong, Singapore, and recently Indonesia) have gone further, restricting entry by new domestic competitors and by foreign banks. As a result few new banks have been established, so the financial sector has expanded largely by licensing new branches of existing banks. Japan's banking system has expanded enormously since the 1950s even though no new banks have been allowed in Japan since then. In Korea and Taiwan (China) only a few new private banks have been licensed in the past few years; competition comes primarily from nonbank financial institutions and the curb market. Not surprisingly, banking in many East Asian economies—Indonesia, Korea, Taiwan (China), and Thailand—has been highly concentrated. Governments had several reasons for imposing entry restrictions.

The most widely cited reasons relate to prudential concerns. Entry restrictions develop from the view that governments must ensure that only trustworthy bankers handle depositors' money. This view is reinforced by the not entirely valid belief that financial systems with a small number of large banks are less risky than systems with a large number of small banks (Vittas 1991). Many policymakers fear that excessively competitive systems—with low profit rates—

Table 1. Nature, causes, and [illegible]

Economy	Nature of financial distress	Causes	Nature of bailout or rescue
Hong Kong 1982–83	Nineteen deposit-taking companies failed.	• Large exposure to real estate lending, fraud and mismanagement, and weak prudential regulation.	• The government revamped regulatory and auditing system and liquidated troubled deposit-taking companies.
Hong Kong 1983–86	Four banks, including a major international bank, became insolvent	• High international interest rates • Large exposure to real estate lending and spillover effects from 1983 crisis because these banks owned deposit-taking companies as subsidiaries	• The government took over the larger banks and introduced new management, including top executives seconded from the largest commercial bank. Those banks not taken over by the government received credit from other commercial banks.
Malaysia 1985–88	The failure of one deposit-taking cooperative in [illegible] caused runs on 32 of 35 others. In addition, 4 of 38 banks and 4 of 47 finance companies were also in financial distress. Over [illegible] percent of banking system deposits were affected.	• Excessive speculation in real estate and stocks • Deterioration in terms of trade	• The government rescued 24 insolvent cooperatives and consolidated and merged weak finance companies. The central bank injected fresh equity capital and replaced management of some banks.
Thailand 1983–87	Government [illegible] 50 finance companies [illegible] was estimated at US$190 [illegible] 48 [illegible]	• Excessive speculation on real estate and exchange rate transactions	• The government liquidated 24 finance companies and merged another 9, and the central bank took over the other 17 and sold them to new investors (including other banks)

	Five commercial banks accounting for [illegible]4 percent of commercial bank assets were in financial difficulties in 1986–87	• High concentration of unsecured insider loans. • High international interest rates	• The government bought some shares of troubled banks. • To provide emergency loans to troubled banks, the government created a "lifeboat fund" financed by contributions from commercial banks.
Taiwan (China) 1983–84	Four trust companies and 11 cooperatives failed	• Cooperatives arbitraging from an artificially steep yield curve	• Healthier banks took over management or bought the shares of failed banks.
Singapore 1982	Domestic commercial banks' nonperforming loans rose to about US$200 million, or 0.6[illegible] percent of GDP	• Macroeconomic reasons.	• The government worked out a two-year write-off period (using tax breaks).
Japan 1991	A central bank report estimated the size of problem loans of the top 21 banks to be between 3.5 percent and 4.8 percent of banking system assets. Informal estimates of the amount to be written off are as much as 1.5 percent of banking system assets	• Excessive exposure to real estate lending (90 percent of bad loans), and a steep decline in real estate prices. • Inadequate prudential supervision. Banks were able to increase their exposure through loans to their nonbank affiliates.	• The government encouraged mergers of weaker banks with healthier ones. • Groups of banks provided emergency loans to weaker banks. • Nonperforming loans were to be transferred to a separate financial institution, and the cost of the write-offs was to be shared among commercial banks.

Source: World Bank (1993a)

Governments set up long-term credit institutions rather than lending or investing directly in firms largely because they believed that a certain amount of independence would enhance the performance of banks and firms. Long-term investments require selection and monitoring. Government agencies are not designed to screen and monitor commercial projects and might be subject to political influence. The creation of the Industrial Bank of Japan reflected a recognition that long-term credit banks had served as effective monitors of firms in other countries, especially among firms not affiliated with major conglomerates (Packer 1992).

Relationship of Development Banks with Government

Governments did more than create the development banks. They also provided assistance, particularly in developing sources of funds during the banks' early years. For example, the Japanese government initially bought a substantial share of the bonds issued by private long-term credit banks and was a catalyst in ensuring that other private banks and financial institutions subscribed to these bonds. It allowed development banks to issue long-term bonds or debentures, whose market the government helped create. This privilege helped redress the mismatch between the maturity structure of the banks' assets and their liabilities, a problem that had plagued commercial banks. Limiting competition enabled the long-term credit banks to obtain funds more cheaply than they otherwise could have. The government went even further, encouraging government units and commercial banks to purchase the long-term bonds, which allowed the development banks to obtain funds at an even lower rate. The Thai government provided similar privileges to its private long term credit banks.

Why do governments form both public and private development banks? The main advantage of private development banks is that they are further removed from government, although the government can still exercise considerable influence. The Industrial Bank of Japan, for example, could choose projects according to its own commercial criteria, but it had to select firms from within priority industries identified by the government. There are tradeoffs: the closer the link between banks and the government, the easier it is for the government to exercise influence and the less likely it is that commercial criteria will be employed. Private development banks may be more credible in setting commercial criteria for investment projects, but private ownership places an additional monitoring burden on government to guard against misappropriation of government assistance. Japan's private and public development banks illustrate how lending activities are divided. The privately owned Industrial Bank is a major lender of long-term funds to industries that are not necessarily being promoted by government, and it has acquired a substantial reputation within these industries. The government-owned Japan Development Bank, on the other hand, focuses its lending on industrial activities that have received the highest subsidies, such as sea transport, mining, electric power, and transport machinery.

The Influence of Development Banks

Development banks lent substantial funds on a long-term basis in Indonesia, Japan, Korea, and Taiwan (China), but not in Hong Kong, which has no development banks, or Thailand, where industrial development banks hold only 1 percent of the assets of the financial system. The Korean Development Bank accounted for about a third of all loans and guarantees in the 1970s. Taiwan's Bank of Communications accounts for about half the assets of the banking system. In Japan the development banks together accounted for about two-thirds of outstanding loans for equipment investment in the 1950s and about half in the early 1960s. The Japan Development Bank alone accounted for 45 percent of equipment lending in the early 1950s. Since the 1950s, however, the Japan Development Bank's lending has accounted for only 2 percent of newly lent funds (Kawaura 1991; Horiuchi and Sui 1993). In recent years the share of development lending has been small even in new growth industries.

DEVELOPMENT LENDING AS A COMPLEMENT TO COMMERCIAL LENDING. For several reasons, however, development banks have been more influential than their small share in lending might suggest. Because they had close ties to the government, their lending provided information to entrepreneurs and other banks on the areas that the government was promoting. In addition, other financial institutions valued information on the development banks' choice of clients (as distinct from sectors). This signaling effect works, of course, only if development banks have sound institutional reputations, which they did in Japan, Singapore, Taiwan (China), and, by and large, Korea.

Development banks did more than provide information. They also initiated loan syndication and other forms of cooperative lending with private banks. Development bank lending often created the perception of "risk sharing"—government support made it more likely that government assistance would be forthcoming in the event of a problem (Yasuda 1992). The perception of implicit government insurance of priority activities was fortified by several government bailouts within priority sectors. At the government's urging, the Industrial Bank of Japan led syndications for firms in distress and forced their restructuring to avoid liquidation or external takeovers. In Indonesia the government bought 35 percent of the equity of a large cement plant, a priority activity, when the plant became financially troubled (Kunio 1988). To protect firms and banks, Korea extended cheap loans during the 1980s to firms that were unable to meet their debt obligations.

Thus development lending complemented private sector lending rather than displaced it. To see how the Japan Development Bank's lending to medium-size and large firms affected corporate investment, Horiuchi and Sui (1993) analyzed a sample of 477 medium-size firms listed in the Tokyo Exchange in 1965. They found that an increase in a Japan Development Bank loan to a firm induced more than an equal increase in the firm's investment spending. Moreover,

private banks extended credit more actively to firms that borrowed from the development bank than to those that did not. Japan Development Bank loans preceded the increase in a firm's borrowing from private banks, further evidence of the signaling effect. Japan Development Bank borrowers also enjoyed more favorable borrowing conditions from other banks and were less sensitive to changes in the cost of capital than were firms that did not borrow from Japan Development Bank.

Even private long-term credit banks seemed to follow the lead of the public development banks. Of 161 listed companies that had a private long-term credit bank as its primary lender in 1967, nearly half also obtained loans from the Japan Development Bank, and 20 percent obtained loans from the Export Import Bank of Japan.

Establishing a causal link between development bank and commercial bank lending is difficult, but several pieces of evidence support the existence of such a link, beyond the evidence on timing noted earlier. Incentives for cooperative financing between Japanese development banks and commercial banks were strong, for example. Commercial banks that cofinanced projects initiated by the long-term credit banks received preferential treatment in purchasing the financial debentures issued by the long-term credit banks.

That development banks influenced lending is important, of course, only if their lending patterns differed from those that commercial banks would have chosen on their own. There is evidence that this is the case. As noted earlier, the discrepancies between private and social returns may be large. Even if such discrepancies did not exist, however, governments believed that they did and expected substantial gains from channeling resources to priority areas where social returns exceeded private returns. Development banks were one of their main tools for channeling such resources. Priority sectors varied across countries and over time. Most countries gave some priority to exports; the Korean Development Bank was a major financial intermediary for loans to heavy and chemical industries in the 1970s, and both the Bank of Communications in Taiwan (China) and Singapore's Development Bank have been active in financing high technology.

SPECIALIZED DEVELOPMENT BANKS. Most East Asian governments have also created specialized banks in areas where private lending has been viewed as inadequate, most notably in agriculture and small scale enterprises. Thailand's agricultural development bank, for example, caters to small farmers who do not have access to commercial bank lending. The bank has reached 80 percent of potential agricultural borrowers (including most small farmers), even though its lending to agriculture is much smaller than total commercial bank lending. Loan rates are slightly lower than commercial bank rates and substantially lower than informal market rates. The bank operates on a cost plus basis (over subsidized funds) and has prudently reduced lending in response to nonrepayment. The bank has been financially sound despite high operating costs and the expected

poor repayment record. To increase farmers' access to formal credit, the Thai government has complemented financial reforms with legal reforms that enable small farmers to use their land as collateral for loans.

Because markets for mortgages—especially for low-income housing—are underdeveloped, East Asian governments have also created financial institutions for housing finance. Japan's postwar government created the Housing Corporation, whose lending has accounted for a growing share of the government's fiscal investment and lending program. Singapore is probably the most prominent example of intervention in housing finance. In 1960 the government established the Housing Development Board to build and provide subsidized mass housing, which policymakers viewed as essential to maintaining social stability. The government subsequently allowed would-be buyers to use part of their provident fund contributions to purchase the subsidized housing. The housing subsidies may have resulted in lower wages (particularly in a country such as Singapore, where wages are determined through national bargaining), but the lower wages boosted profits and, consequently, retained earnings. Moreover, the housing program increased social stability by raising living standards and increasing the net worth of households, giving them a greater financial stake in their society.

Why Have Development Banks Succeeded in East Asia?

Many developing countries have been unsuccessful in promoting development banks. Eighteen development banks examined in a World Bank (1989) study had, on average, half of their loans in arrears. Even in East Asia, where the experience with development banks has been mostly positive, development banks have failed. Japan's insolvent Reconstruction Finance Bureau was closed in 1952 and Thailand's Industrial Bank in 1959. Some development banks in Indonesia and Malaysia are reporting a rising volume of arrears. The most common causes of failure are political pressure to finance bad projects and poor incentives for financial institutions to screen and monitor projects.

Among the key ingredients of the success of many other East Asian development banks was an insistence on commercial standards within the priority sectors. Successful development banks transformed themselves from government agencies financing development projects into more market-oriented financial enterprises. The largest development banks of Japan, Korea, Singapore, and Taiwan (China) have consistently demonstrated such a pattern.

Although government monitoring seems to have had a salutary effect in East Asia (Cho and Hellman 1993), government intervention has had a negative impact in other countries. This differentiated outcome is likely the result of government efforts to shield development banks from political interference. Japan, Korea, and Taiwan (China) appointed established officials (from ministries of finance) as chairmen so that they could better withstand pressure from other parts of the government. That does not explain why these officials did not subvert the development banks for their own purposes, however. And the fact that

these officials were competent and honest does not mean that in other countries and at other times oversight by a ministry of finance will provide an adequate check against abuse. Some East Asian countries have also controlled types of lending: Thailand simply barred its development bank from lending to state enterprises. Successful development banks also instilled a high level of professionalism and institutional identification in their staff, making government intervention—other than in establishing priority sectors—difficult. By making lending to a nonperformer a criminal offense, Taiwan (China) made sure that loan officers did not give in to political pressures or abuse their discretion. Moreover, private banks often cofinanced development bank projects, thereby serving as a check on development banks' lending criteria. Thus the information flow between development banks and commercial banks was reciprocal.

Creating Financial Markets

Only a small portion of long-term investments in East Asia have been financed by corporate bonds. Except for Thailand and Korea since 1980, bonds accounted for much less than 10 percent of the net financing of nonfinancial corporations among five high-performing East Asian economies for which data exist (World Bank 1993a). One obstacle to the emergence of bond markets in these economies is the absence of a market for government securities—because the governments do not run deficits, they do not need to borrow. With no market for government securities, there is no benchmark risk-free rate, and so markets must determine both the risk-free rate and the risk premium associated with a specific corporate bond. Hong Kong's government has responded to this limitation by auctioning government bonds—even though it does not need the financing—to provide a benchmark risk-free rate and eventually help create a market for corporate bonds. Malaysia and Singapore are considering doing the same.

Other East Asian economies have also taken steps to foster the growth of bond markets. Malaysia, for example, established a rating agency for bond issues in 1991. Hong Kong, Taiwan (China), and Thailand have strengthened their legal infrastructure for securities (bonds and equity) issues. And in Korea—which has the region's most rapidly expanding bond market—the government has been issuing guarantees.

Like bond markets, equity markets provide a small fraction of the net financing of nonfinancial corporations in East Asia, although the relative importance of equities rose slightly in Korea and Thailand during the 1980s. In recent years East Asian governments have increased their efforts to promote stock markets. Hong Kong has strengthened disclosure requirements and implemented laws against fraud in response to financial disruptions experienced in the stock exchange. Korea, Taiwan (China), and Thailand have provided preferential corporate tax measures to encourage companies to list on the stock exchange. Korea introduced these incentives in the 1970s and since the 1980s has indirectly

promoted equity issues by encouraging firms to lower their debt-equity ratios. All of these East Asian economies have expanded their stock exchanges; Indonesia, Korea, and Singapore have introduced over-the-counter markets.

Enhancing Growth through Financial Restraint and Credit Allocation

Earlier we showed how, contrary to standard arguments, financial restraint may have increased national savings. A standard argument against financial restraint has been that it impedes efficient resource allocation by preventing a free market auction from occurring (World Bank 1989; Fry 1988). As recent work on credit markets has emphasized, however, credit is not allocated by auction even in perfectly competitive markets. In a world of asymmetric information, banks do not allocate loans to the highest bidders, but rather to those borrowers they deem most likely to repay. Even when the adverse selection and incentive effects associated with higher interest rates do not induce credit rationing (Stiglitz and Weiss 1981), these effects do mean that moderate financial restraint on lending rates reduces default rates and increases the social returns to lending.

Financial restraint has further allocative benefits. To the extent that lower deposit rates are reflected in lower lending rates, financial restraint enhances the ability of firms to increase their equity, and hence their level of investment and their ability and willingness to take prudent risks (Stiglitz 1994). To the extent that lower deposit rates are not passed on in lower lending rates, financial restraint enhances bank equity, and hence banks' ability and willingness to make loans. And greater bank equity enhances the stability of the financial system. One of the benefits of stable financial systems is the organization-specific nature of information: bank failures destroy information that is valuable to ensuring the efficient allocation of capital.

Financial restraint also has incentive effects. Higher bank profits increase the franchise value of banks, providing strong incentives for prudential behavior. Appropriately chosen bases for allocating scarce credit can also provide strong performance incentives. In Japan, Korea, and Taiwan (China), competition for access to credit generated high marginal returns to greater effort, as measured, for instance, by exports (Stiglitz 1994). Because the shadow value of access to capital was high, this prize was valuable.

The empirical evidence usually cited against financial restraint, based on cross-country regressions showing a positive relation between real interest rates and output growth (Gelb 1989), is as faulty as the theoretical arguments. This evidence suggests that high real interest rates are associated with increased financial depth, a modest increase in savings and investment, and more productive investments (World Bank 1989). That savings and income have continued to grow in East Asian countries despite financial restraint raises some questions

about the findings of these studies: does anyone really believe growth would have been even faster in East Asia in the absence of financial restraint?

There are three problems with these studies (see Stiglitz 1994 and Murdock and Stiglitz 1993 for greater detail). First, they fail to distinguish between large and moderate degrees of interest rate constraints. Highly negative real rates have a severely negative impact on economic performance, which the regressions capture, but moderate interest rate restraints may yield positive effects, which the regressions do not reflect. Indeed, when the sample is split between developing countries with high and those with moderate degrees of financial repression, the positive correlation between real interest rates and growth disappears for the countries with moderate financial restraint.

Second, countries with severely negative real interest rates have had both severe financial repression and bad macroeconomic policies (as reflected in high rates of inflation). These policies, rather than the financial restraint, may account for their poor economic performance. That suggests that the gains to growth come less from rationalizing interest rates and more from decreasing the distortionary effects of high inflation.

The third problem is one of identification: real interest rates may be low not because of financial restraint, but because there are no good investment opportunities. This problem can be addressed in two ways: using a simultaneous equations model or trying to measure financial restraint directly. Such assessments are possible for Korea and Taiwan (China), using data on curb market rates. For these cases the degree of financial restraint does not appear to explain economic growth, although other variables, such as inflation, do.

Priorities Reflected through Directed Credit

All East Asian countries have directed credit in varying degrees to support industrial policies or social objectives. Countries direct credit for several reasons, ranging from a perceived contrast between social and private economic rates of return to more immediate concerns, such as providing national security. Like other economies, high-performing East Asian economies use two broad types of intervention. First, the government directs credit to priority firms, groups, industries, and activities (such as exports or high-technology projects). Second, the government directs credit for social reasons, often to small farmers, small and medium-scale enterprises, or a specific ethnic group. In both cases the government directs credit by investing in public enterprises, using its development banks to lend to priority areas (and to signal to other financial institutions what these areas are), and compelling commercial banks to lend to designated activities. Although the rationale for and categories of directed credit do not differ between the East Asian economies and other economies that have used directed credit, the high performers in East Asia have implemented their programs with more moderate credit subsidies and with institutions that enabled better selec-

tion (through performance criteria) and monitoring of promoted projects, resulting in higher repayment rates for subsidized loans.

Among East Asian economies Japan and Korea most pervasively directed credit to promote specific firms and industries—Japan during its postwar reconstruction and Korea during its promotion of chemical and heavy industries in the 1970s. During the 1950s the Japanese government's financing amounted to nearly a third of new equipment lending to industry; most went to shipbuilding, electrical power, coal, sea transport, and machinery. The results of the program in Japan are controversial: many successful growth industries were not heavily supported by the government's credit programs; among those provided with credit subsidies, some (such as shipbuilding) increased their exports, while others (such as coal mining) continued to decline. The results of Koreas's policy loans also have been mixed (Stern and others 1992). Some of Korea's heavy industries, such as steel, electronics, and passenger cars, became leading exporters during the 1980s, while others became financially distressed. The chemical and heavy industries policies also increased the concentration of wealth among conglomerates and contributed to high firm leverage (Cho and Kim 1995).

Indonesia and Malaysia had few successful experiences with selective credit intervention and abandoned the schemes once the negative effects of the policy became apparent. Thailand has avoided credit programs directed at specific firms and industries.

Assessing the success or failure of a directed credit program is difficult for three reasons. First, as mentioned earlier, there is usually no way of knowing whether growth would have been higher or lower in the absence of the directed credit program. Second, a good program requires risk taking, which means that failures are inevitable. A program with nothing but successes would necessarily have been too conservative. Third, many of the returns may be long term, so current profitability may not provide an adequate measure of success. Thus the measure of Korea's chemical and heavy industry program should not be how those industries fared in the early 1980s, but what the structure of the economy looks like in the late 1990s. By the same token, low profits may reflect cyclical conditions rather than long run prospects.

Thus, although it is difficult to determine whether directed credit programs were successful, the evidence shows that government lending did not simply displace private lending—it affected the allocation of resources. Moreover, credit was directed to areas with high social returns. During the 1950s in Japan, for instance, the bulk of directed credit went to basic industries that supplied essential inputs for growth in other parts of the economy. Once these basic industries were developed, Japan promoted other industries (such as machine tools) whose expected spillover effects on the economy were large (JDB/JERI 1993).

Many other developing countries have failed in their industrial credit programs. At least six factors set the successful East Asian economies apart from others. First was their ability to change credit policies rapidly when they real-

ized the policies were not functioning properly. Second, unlike many developing countries that funneled a large portion of directed credit to public enterprises, the high-performing East Asian economies directed credit mainly to private enterprises. Even in East Asian countries that made loans to state-owned firms—Indonesia, Malaysia, and Singapore—the proportions of total credit were not persistently high, the state firms tended to perform better financially than state firms in other countries, and the interest rate subsidies were not substantial (except in Indonesia). Third, all East Asian economies directed credit to industry based on broad functional criteria (such as whether the firm produced exports), typically using objective performance measures. Fourth, credit was usually more common than outright subsidies, which were limited. Fifth, directed credit was more limited than elsewhere. Although directed credit amounts to as much as 75 percent of the loans of financial institutions in some countries, even in Korea (which used directed credit most aggressively) directed credit amounted to only about 40 percent of total credit, and in Japan it never exceeded 15 percent. Sixth, monitoring was more effective, so default rates were lower.

Can the East Asian Experience Be Replicated?

Most East Asian governments' financial sector interventions were meant to remedy market failures. Such failures occur even in industrial countries, and governments impose extensive regulation to deal with them. Market failures are usually more significant in developing countries, and governments' ability to correct them is more circumscribed. What is remarkable is that East Asian governments undertook actions (such as prudential regulation) similar to those taken by more industrial countries, and that they did so at an earlier stage of development. Moreover, these regulatory initiatives succeeded without the abuses that often accompany them elsewhere. East Asian governments sought not to replace markets and market forces, but to use and direct them. Government lending programs complemented private lending; they did not replace or displace it. Although governments established priorities for lending—and discouraged lending for real estate and consumer goods—they still employed commercial standards.

How replicable are these interventions in other developing countries? That so many East Asian economies were successful suggests that success was not just fortuitous, the result of, say, unusually good civil servants. The cultural diversity of the region makes explanations based on unique cultural factors unpersuasive (Stiglitz 1994).

Many of the specific institutions, programs, and practices that contributed to East Asia's success can easily be replicated, including the region's large investment in education. The resultant high level of educational attainment undoubtedly contributed to governments' ability to execute their programs. Several of the institutions that contributed to high savings—such as the postal saving system and provident funds—could easily be introduced elsewhere. Prudential regu-

lations, particularly capital adequacy requirements and controls on real estate lending, are essential and replicable. The adaptability of government policies—the ability to abandon policies when they fail and to change policies with changing circumstances—is clearly a lesson of general applicability, although it is hard to design institutions that capture that lesson.

The main concern with activist policies such as those pursued in East Asia is abuse of political power: activist policies generally entail giving governments discretionary powers that can easily be abused. In other countries such abuse explains or at least contributes to the failure of activist programs. But many of the ways in which East Asian institutions were designed reduced their vulnerability to political abuse—and these institutional arrangements can be replicated. The use of performance-based criteria for allocating credit, for example, limited discretion. Similarly, requiring that commercial criteria be satisfied to receive credit and requiring borrowers to raise part of their funds on their own and to put up their own equity are replicable practices that enhance the likelihood that funds will be allocated to good ventures and reduce the likelihood of political abuse.

Furthermore, moderate subsidies and financial restraint reduced rents (relative to those found in many developing countries), further attenuating incentives for abuse. So did the extensive use of competition (Stiglitz 1994). Japan used several tools, including interest rate regulation and competition from nonbank financial institutions, to curtail potentially high rents from entry restrictions.

At bottom, however, is a more fundamental issue: why were the governments of East Asia able to implement policies that lessened the potential for abuse? The answer lies in political economy, which is beyond the scope of this article. A few studies have touched on this issue (Campos 1993), but much remains to be done.

Notes

Joseph E. Stiglitz is chairman of President Clinton's Council of Economic Advisers, on leave from Stanford University, where he is professor of economics, and Marilou Uy is with the South Asia Country Department of the World Bank. This article was written as part of the World Bank's project on the East Asian miracle. The views expressed in it are entirely those of the authors and in no way represent those of the organizations with which they are affiliated. The authors gratefully acknowledge the research assistance of Kevin Murdock, as well as the comments of participants at the many seminars at which portions of this article were presented. The authors also acknowledge the assistance of Brian [illegible] in preparing the figures.

1. Many of the theoretical and technical details of the analysis presented in this article, especially on the market failure framework, are discussed in Stiglitz (1985, 1989, 1994, and 1996) and Greenwald and Stiglitz (1986).

2. Standard economic theories have identified several reasons why savings rates might depend in part on income growth rates (Carroll, Weil, and Summers 1993), although the sign of the relationship is ambiguous. First, as growth rates increase, households become less confident that such growth will be sustained; it becomes more probable that growth

will fall, and so individuals are induced to put away more for a "rainy day." Second, as growth rates increase, the degree of uncertainty increases, which may induce households to save more. Third, at higher growth rates, the return to capital may be higher, and higher returns to capital may elicit higher savings rates. And fourth, lags in adjustment (habit formation) may result in consumption not keeping up with income. That is perhaps the most persuasive reason for the high savings rate. Other arguments suggest that faster growth rates would be associated with lower savings rates. With higher growth rates, individuals need to save less for the future.

3. To be sure, in a world with perfect information, depositors would recognize the risks associated with undercapitalized and low-profit banks and would deposit their money with such institutions only if the higher rate of interest paid on deposits compensated for the risk. These higher required deposit rates would serve as a barrier to entry, ensuring that even without government intervention there would not be "excessive" entry; evidence for this exists for some countries. Still, in many of the countries of East Asia, it appears that—at least at their stage of development—there would have been far more entry in the absence of government intervention.

4. Even with entry restrictions, competition remained keen in most East Asian countries (for Japan, see Sakakibara, Feldman, and Harada 1982). Japanese banks have lower margins than banks in the United States (Ueda 1992) or Germany. Gross interest markups as well as net operating costs (as a percentage of total assets) of banks in Korea, Malaysia, and Thailand were in line with those in industrial countries in the early 1980s and lower than in many other developing countries, notably Turkey and selected countries in Latin America (Hanson and Rocha 1986). On the other hand, World Bank studies (World Bank 1993b) of the banking systems in Indonesia and Korea find that protection has led to less efficient and innovative practices. Moreover, even in Japan there has been less innovation than in the United States, for example, in introducing derivative securities.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Balassa, Bela. 1989. "The Effects of Interest Rates on Savings in Developing Countries." Policy Research Working Paper 56. World Bank, Policy Research Department, Washington, D.C. Processed.

Campos, Jose Edgardo. 1993. "The Institutional Foundations of High Speed Growth in the High-Performing Asian Economies: Insulation Mechanisms and Public Sector Private Sector Relations." Background paper for *The East Asian Miracle*. World Bank Policy Research Department, Washington, D.C. Processed.

Carroll, Chris, David N. Weil, and Lawrence H. Summers. 1993. "Savings and Growth." Paper presented at the Bradley Policy Research Center's Carnegie Rochester Public Policy Conference, Rochester, N.Y. Processed.

Chiu, Paul C. H. 1992. "Money and Financial Markets: The Domestic Perspective." In Gustav Ranis, ed., *Taiwan from Developing to Mature Economy*. Boulder, Colo.: Westview Press.

Cho, Yoon Jee, and T. Hellman. 1993. "Government Intervention in Credit Markets: An Alternative Interpretation of Japanese and Korean Experiences from the New Institutional Economics Perspective." World Bank, Financial Sector Department, Washington, D.C. Processed.

Cho, Yoon Jee, and Joon-Kyung Kim. 1995. *Credit Policies and the Industrialization of Korea*. World Bank Discussion Paper 286. Washington, D.C.

Corbo, Vittorio, and Klaus Schmidt-Hebbel. 1991. "Public Policy and Savings in Developing Countries." *Journal of Development Economics* 36(1):89–115.

Development: A Comparative Study of Selected Asian and Latin American Countries, Honolulu. October. Processed.

Noguchi, Y. 1985. "Pensions, the Tax System and the Saving Behavior of the People." In *Kenkyu Hopoku Gaiyo [Summary of Research Findings]*. Research Project on the Future Prospects of Government-Managed Voluntary Life Insurance, Tokyo. (In Japanese.)

Packer, Frank. 1992. "The Role of Long-Term Credit Banks within the Japanese Main Bank System." Columbia University, New York. Processed.

Sakakibara, E., R. Feldman, and Y. Harada. 1982. *The Japanese Financial System in Comparative Perspective*. Washington, D.C.: U.S. Government Printing Office.

Shea, J. D. 1993. "Development and Structural Change of the Financial System." In Patrick Hugh and Yung Chul Park, eds., *Financial Development—Japan, Korea and Taiwan* New York: Oxford University Press.

Singapore Department of Statistics. Various years *Yearbook of Statistics* Singapore

Stern, Joseph, Ji-Hong Kim, Dwight Perkins, and Jung-Ho Yoo. 1992 "Industrialization and the State: The Korean Heavy and Chemical Industry Drive" Harvard Institute for International Development, Cambridge, Mass., and Korea Development Institute, Seoul Processed.

Stiglitz, Joseph E. 1985. "Credit Markets and the Control of Capital" *Journal of Money Credit, and Banking* 17:133–52.

———. 1989. "The Economic Role of the State" In Arnold Heertje, ed , *The Economic Role of the State*. Oxford, U.K.: Basil Blackwell.

———. 1994. "The Role of the State in Financial Markets" In Michael Bruno and Boris Pleskovic, eds., *Proceedings of the World Bank Annual Conference on Development Economics 1993*. Washington, D.C.: World Bank

———. 1996. "Some Lessons from the East Asian Miracle" *World Bank Research Observer* 11(2):151–77.

Stiglitz, Joseph E., and A. Weiss. 1981 "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71(3):393–410

Sun, C., and M. Liang. 1982. "Savings in Taiwan, 1953–1980" In Kwoh-Ting Li and Tzong-Shian Yu, eds., *Experiences and Lessons of Economic Development in Taiwan*. Taipei: Institute of Economics, Academia Sinica

Ueda, Kazuo. 1992. *Institutional and Regulatory Frameworks for the Main Bank System* Tokyo. University of Tokyo.

Vittas, Dimitri. 1991. "Measuring Commercial Bank Efficiency Use and Misuse of Bank Operating Ratios." Policy Research Working Paper 806 World Bank, Country Economics Department, Washington, D.C.

World Bank. 1989. *World Development Report 1989. Financial Systems and Development*. New York: Oxford University Press.

———. 1993a. *The East Asian Miracle: Economic Growth and Public Policy* New York Oxford University Press.

——— 1993b. "Korea Financial Sector Study." East Asia and Pacific Regional Office Country Department I, Industry and Energy Operations Division, Washington, D.C. Processed.

Yasuda, Ayako. 1992. "The Performance and Roles of Japanese Development Banks" Stanford University, Directed Readings for Professor Joseph E Stiglitz, Stanford, Calif. Processed.

Yoshino, N. 1992. "History and the Role of Post Office Savings and the Role of Government Financing in Japan." Keio University, Department of Economics, Tokyo Processed.

CREDIT POLICIES: LESSONS FROM JAPAN AND KOREA

Dimitri Vittas
Yoon Je Cho

The success of policy-based credit programs in Japan and the Republic of Korea suggests that credit policy can be an effective instrument for economic development. Why, then, have credit policies failed in so many countries, and what factors explain their relative success in Japan and Korea?

Both economic and institutional factors appear to be important in the success or failure of credit policies. Essential economic factors include a reliance on the private sector, a bias toward industrialization, an orientation toward export production, the encouragement of domestic competition, and a commitment to price stability. Crucial institutional factors include extensive and frequent consultation between government and the private sector, effective monitoring systems, and, most important, a clear and credible plan for economic development. Although several countries have included one or more of these factors in their programs, the experience of Japan and Korea suggests that a comprehensive network combining all or most of these factors may be necessary for the successful implementation of credit policies.

Directed credit programs that give loans on preferential terms and conditions to priority sectors were a leading tool of development policy in the 1960s and 1970s. The realization that most of these programs had created distorted economic incentives among both lenders and borrowers led to a reconsideration of their rationale and effectiveness during the 1980s. Countries around the world found that the programs had stimulated projects that were capital intensive, that preferential funds were sometimes used for nonpriority purposes, and that the programs had increased the cost of funds to nonpreferential borrowers. In addition, policy-based credit programs had provoked a decline in financial discipline that led to low repayment rates and a swelling of budget deficits. Once introduced, moreover, policy-based credit programs proved difficult to eliminate.

The World Bank Research Observer, vol. 11, no. 2 (August 1996), pp. 277–98

This negative assessment conflicts, however, with the experience of directed credit programs in Japan and the Republic of Korea. Government officials in these countries argue that subsidized interest rates and government involvement in directing credit are warranted under several circumstances: when the economy shows a significant discrepancy between private and social benefits; when the investment risk of particular projects is too high to attract private lenders; when the lack of reliable economic information discourages private lending to small and medium-size firms; and when infant industries face large social set-up costs (OECF 1991). They suggest that the main constraint facing new or expanding enterprises is limited access to credit at reasonable terms and conditions and that policy-based lending and other forms of industrial assistance such as grants and tax reduction can overcome this constraint.

Theoretical Underpinnings of Policy-Based Lending

In an ideal world in which economic information is complete and readily available, the financial system is passive. Investors fund the projects that yield the highest returns, and neither governments nor financial institutions need to improve the allocation of credit. In the real world, however, information is highly imperfect and costly to acquire, and the allocation of credit suffers from the unequal distribution of information, the costs of monitoring and verification and the costs of default or contract enforcement. Under these conditions, credit is not necessarily allocated to its best use.

Informational asymmetries give rise to the possibility that credit may be given to unviable candidates (adverse selection), that it may be awarded to irresponsible parties (moral hazard), that some players will attempt to receive, without cost, the benefits of credit allocation (free riding), and that incentives arising from the credit program itself may conflict with one another or with program goals. These problems may be further compounded by uncertainty about projects' returns and by dynamic externalities, which occur when external, or social, benefits increase faster than do private benefits. The potential for difficulties of this kind justifies intervention by governments and financial institutions in the allocation of credit—even though their ability to allocate credit efficiently will also be constrained by the lack of information.

Economic theory has made considerable progress in recent years in understanding these phenomena and has bridged the wide gap that once existed between theory and practice (Calomiris and others 1992; Calomiris and Himmelberg 1994, 1995; Cho and Hellmann 1994). Economic theory now stresses, for example, the role of market imperfections in explaining why firms rely on internally generated funds (retained earnings) and other forms of "inside" finance (finance provided by owners, managers, and banks that have access to information not available to the public). Reliance on inside finance is especially pronounced for young growing firms and for new indus

trial sectors in developing countries, and it is clearly a factor constraining industrial growth.

Lending by Banks

In the absence of full information, banks tend to allocate credit to firms with reliable track records or available internal funds, even if other firms present better investment opportunities. Financial intermediaries, especially (but not only) commercial banks, can acquire information that is superior to that of outsiders by developing and maintaining close long-term relationships with their customers. They can thus play an important role in screening projects, monitoring behavior and outcomes, and managing corporate distress.

These potential advantages of bank lending, however, depend on the behavior of the bankers involved and the incentives and regulations that govern their operations. In many countries, commercial banks favor lending for low-risk activities, such as self liquidating, short-term working capital and trade finance, or for high risk, but more speculative, projects with short payback periods, such as real estate development. That is especially true for countries where information on corporate performance is difficult to obtain, but it is also true for the United Kingdom, the United States, and the Scandinavian countries, which do not suffer from such problems. Commercial banks are generally less willing to finance high risk projects with long payback periods, even if these projects may yield higher overall returns. They are generally also reluctant to finance small firms that lack adequate collateral, even though such firms may be more innovative and promising than others.

Lending by Governments

A government role in allocating credit can be justified on two grounds. First, directed credit programs can be a preferred or superior industrial policy instrument for reaping positive externalities (that is, for increasing benefits across the economy). If firms lack access to credit, other industrial policy tools, such as tariffs and subsidies, that may rely on cost and profit incentives to increase production could prove ineffective.

Second, the government has a comparative advantage in directing the allocation of credit. Government agencies (often in direct collaboration with private industrial associations and research institutes) may have better information on sectoral prospects than do individual private firms. They may therefore have an advantage in screening projects, as well as in monitoring behavior and outcomes—although that will depend partly on the relative efficiency of the government agencies and the financial intermediaries. In addition, government costs of enforcing contracts, through taxation and police powers, are likely to be lower than costs for private intermediaries. The power to tax, moreover, may make it possible for the government to internalize benefits from specific lending policies

that private intermediaries cannot capture. This advantage increases if there are technological spillover effects that neither the firm nor its intermediary can capture but that the government can claim through future taxes.

Ultimately, however, the advantages depend on the motivation and the efficiency of the government involved. Governments do not always "do the right thing." Government involvement in credit allocation can, and often does, result in rent seeking by borrowers, corruption by bankers and government officials, and crowding out of other worthwhile projects. An important issue in the study of policy-based lending is how governments can prevent rent-seeking behavior by borrowers from undermining the growth objectives of government policies.

Experience with directed credit policies varies widely. In Japan and Korea government intervention in credit markets is deemed to have been effective and beneficial for economic growth and development. In the vast majority of developing countries, however, credit policies have failed to promote growth and have given rise, instead, to severe market distortions.

Size and Scope of Programs

Most studies of directed credit policies focus on the size of the programs, the level of interest rates, and, especially, the level of subsidies. But focusing too narrowly on these aspects of credit policy may be misleading. In many cases, the government's influence on credit allocation may be much stronger than these levels indicate. Japan, for example, avoided highly negative real interest rates (rates adjusted for inflation) during its early period of economic development and had relatively small (explicitly labeled) directed credit programs. The government's role in credit allocation was nonetheless significant; in this early period Japan combined policy-based credit with extensive financial regulation to create a rigidly segmented financial system that favored lending to industry but discouraged lending for speculative purposes, real estate development, or consumption (Vittas and Kawaura 1995).

Similarly, one cannot assess the degree of government intervention in a country over time by simply looking at the level of real interest rates. Korea, for example, doubled the level of interest rates in 1965, yielding highly positive real rates. Its action was interpreted by many as financial liberalization, but the interest-rate rise actually strengthened the government's role in allocating credit by shifting funds from the unregulated curb (informal) market to the banks, which were more tightly controlled by the government (Cho and Kim 1995).

Program Management

Just as good or bad management affects the performance of firms, good or bad governance affects the success of credit policies. Although economists have recognized some merit to government intervention in specific markets, they have

often equated failures in these markets with a simplistic notion of government failure. Precisely what characteristics contribute to good economic management and good policy implementation?

Good management, even for an economy, does not always mean the least management or least intervention. It does, however, require effective incentive schemes to motivate behavior beneficial to the economy; it also requires a supportive institutional environment, an arrangement for close consultation and coordination between the government and business sector, and effective mechanisms for monitoring and enforcing performance. Government control of credit may also be understood in light of governance structures. Indeed, credit allocation was used as a powerful instrument for governance of industrial firms in Korea and, to a lesser extent, in Japan (Cho and Hellmann 1994; Vittas and Kawaura 1995).

Government's role in credit policy should also be understood in a dynamic context. In the early stages of economic development, when many markets are missing, when those that exist are highly imperfect, and when private institutions are poorly developed, government intervention can help stimulate healthy economic growth. In the later stages of development, when the private industrial sector has become more sophisticated and markets are better developed and more robust, the merits of government intervention diminish significantly.

Credit Supports or Grants?

Assuming that the government is able to identify and confront the problems arising from market imperfections, why should it use credit supports (preferential access to credit or subsidized interest rates) rather than grants (direct outlays through the government budget) to address these problems? The answer relates to both efficiency and flexibility. In disbursing grants, the government must select the individual firms that deserve support. In disbursing credit subsidies, the government need only signal the sectors that deserve support and leave the selection of individual firms to the banks. In addition, directed credit as an instrument of industrial policy depends partly on the timing of the subsidy, not simply the amount. Subsidies may be allocated more flexibly than grants, and they may take into account the performance of subsidized firms or industries. Good performance can be rewarded by rolling over debt or extending new debt; bad performance (including the diversion of funds to nonpriority uses) can be punished by reducing or terminating support. In addition, it is not clear, despite claims to the contrary, that budgetary outlays are more open to scrutiny than are credit subsidies. Both kinds of support risk politicization. In a country where individual credit allocations are based on economic criteria, operating through credit subsidies may be both less politicized and more flexible than operating through the budget.

Credit Policy in Japan

Japanese credit and industrial policy has evolved in response to the changing needs and structure of Japan's economy (Vittas and Wang 1991; JDB/JERI 1994; Vittas and Kawaura 1995). Three phases are usually identified: the reconstruction period, from 1945 to 1955, when industrial policy and the direct government allocation of funds were most significant; the high-growth period, from 1955 to 1973, when government policy operated less directly, although the financial system was rigidly segmented and subject to wide-ranging controls; and the liberalization period, from the mid-1970s to the present, when policy became less interventionist, and a slow but steady process of financial liberalization began.

Japan's early emphasis on individual industries has given way to a recent concentration on programs that cross industrial boundaries. The main focus of industrial policy in the 1950s was the expansion of capacity, followed in the 1960s by modernization and technological upgrading, in the 1970s by restructuring and adjustment at both the company and industry levels, and in the 1980s by diversification of the industrial structure. Throughout, Japanese credit (and industrial) policy seems to have had four specific industrial objectives: to pick and support "winning" industries, especially in markets in which Japan could enjoy a dynamic comparative advantage; to phase out industries in which Japan was no longer internationally competitive; to support small firms; and to provide the industrial infrastructure necessary for growth.

The extensive financial support given some traditional and relatively inefficient industries, such as agriculture, can be seen more as a social policy objective, or a response to political pressures, than as a component of an active industrial policy. Government regulation and intervention in the financial sector did not focus exclusively on securing cheap finance for the most dynamic sectors but instead seemed to aim for a balance among the claims of different sectors

Basic Features

During the high-growth era, several features of the Japanese financial system, although not unique in themselves, combined to produce a system quite distinct from Anglo-American or European financial systems. These features included the preponderant use of indirect finance; the reliance of the large city banks on credits from the Bank of Japan to fund their loans to industrial corporations; the overborrowing, or high leverage, of industrial companies; and the artificially low level of interest rates (Suzuki 1980; but see Kuroda and Oritani 1980, Horiuchi 1984, and Ikeo 1987, who challenge the particularity of these characteristics). Other distinctive features of the Japanese financial system include the role played by the main bank system (in which certain banks were not only main lenders to specific companies but also the main shareholders and providers of other financial services), the close relations between banks and industry, the

different roles played by debt and equity in the Japanese financial system, and the important financial intermediary role played by large conglomerate groups, especially the general trading companies, in channeling funds to small firms at the periphery of different industrial groups. Because the financial sector in Japan was highly segmented during this period, the government had considerable control over the allocation of financial resources. The Japanese authorities did not, however, impose strict directed credit programs on private financial institutions.

The most distinctive element of Japan's industrial policy has been the cooperative relationship between government and industry, each of which has recognized the necessity of striving toward common goals. This cooperation has been reflected most clearly in the emergence of the "deliberative council system," by which councils made up of industry representatives, former bureaucrats, academics, and others have provided the public and private sectors with a forum for coordinating and developing policy directives. These councils were well established by the 1960s, when they decided almost all important industrial policies.

Another important aspect of Japan's directed credit programs has been the high quality of loan appraisal and project oversight. Loan approval is based on detailed reviews of the projects to be financed and evaluations of the history and character of the firms involved. Once projects are approved, no payments are made without adequate documentation. Close cooperation between development and commercial banks then ensures continuous monitoring of the performance of borrowers and allows development banks to take early action if loan repayments are in arrears. The general economic success of Japan has also meant that most borrowers have made substantial profits and have had little difficulty in repaying their loans.

Three general characteristics of Japanese credit policy seem particularly significant in accounting for its success. The first is the Japanese government's *respect for the market economy*. A precondition for a successful policy-based financial system is the existence of private business and financial structures that can be supplemented with policy-based funding. Japan's prewar experience as a market economy and its determination to establish postwar economic reform created an environment that encouraged entrepreneurship within the existing private sector.

The second characteristic (particularly evident as Japan's high growth period unfolded) is the *close relation between policy-based finance and the government's economic plans*. The preconditions for this relationship are a public savings system (in Japan, the postal savings system) and a vehicle for the efficient allocation of funds (in Japan, the Fiscal Investment and Loan Program).

The third characteristic has been the Japanese government's *respect for managerial independence*, assuming the existence of a sound financial institution and management. Although the Japan Development Bank has had inherent limitations as a government-related financial institution, it has the autonomy to make funding decisions on a neutral and fair appraisal basis.

The Reconstruction Finance Bank and the Dodge Plan

The Japanese government established several agencies to assist in the distribution of policy-based funds. The first of these was the Reconstruction Finance Bank, which was created immediately after the war to provide finance to priority industries. It accounted for 84 percent of the total funding for capital investment during the postwar period and for 16 percent of the working capital needs of major industries such as coal, electric power, fertilizers, iron and steel, ocean shipping, and textiles. This financing facilitated the recovery of production of these high-priority industries and paved the way for Japan's economic recovery.

The Reconstruction Finance Bank was funded, however, by bonds underwritten by the Bank of Japan, a strategy that fueled inflation, which in turn made the fixed interest rates on priority loans highly negative. The loss of control over inflation, combined with a high incidence of delinquent loans (as well as financial scandal at the reconstruction bank), led the Japanese authorities to suspend the bank's operations in 1949, at which time the Dodge Plan for economic stabilization was introduced.

The Dodge Plan was intended to achieve a central government surplus and a unified and stable exchange rate. The plan initially caused a deep economic recession, with reductions in exports and investment, company closures, and production and personnel cuts. Successful implementation of the Dodge Plan, however, imposed fiscal balance, removed price controls, and contained monetary expansion. Although the policy brought about deflation, it created the stable macroeconomic environment necessary for the subsequent implementation of policy-based lending and other industrial policies leading to the rationalization and modernization of Japanese industry.

Other Directed Credit Institutions

In the early 1950s the Japanese authorities established several policy-based financial institutions to provide funding for industrial investment, housing development, and other purposes.

THE FISCAL INVESTMENT AND LOAN PROGRAM. To avoid the inflationary implications of financing these activities through monetary creation, these institutions were funded with resources collected through the extensive postal savings and annuities network and channeled through the Trust Fund Bureau as part of the Fiscal Investment and Loan Program. The funds available to the program amounted to 4 percent of gross national product in the 1950s and increased to more than 8 percent in the early 1990s. This increase reflects the growing importance of the postal savings funds. Policy-based finance through the trust fund accounted for 13 percent of total lending in the mid-1950s, fell to 10 percent in the 1960s, rose to 15 percent in the 1970s and 1980s, and declined to 12 percent in the early 1990s.

The funds have increasingly been used over the years for social (or, at least, nonindustrial) purposes, especially for financing housing development. The trust fund's share of the total supply of funds for new industrial equipment fell from about 30 percent in the mid-1950s, to 20 percent for most of the following two decades, to about 12 percent in recent years.

The most favorable interest rate offered by policy-based financial institutions at the beginning of the 1950s was 3.5 percentage points lower than the private sector long-term prime rate; the least favorable was the same as the prime rate. In addition, policy-based loans had much longer maturities (up to twelve years) and did not require the compensating balances that often substantially increased the cost of private finance, especially for smaller firms. Although policy-based loans continued at fixed rates of interest after the 1950s, success in maintaining macroeconomic and price stability avoided the recurrence of the highly negative interest rates that had bedeviled the reconstruction bank's early operations.

THE JAPAN DEVELOPMENT BANK. The successor, in a sense, to the Reconstruction Finance Bank, the Japan Development Bank was given a managerial independence the first bank had never had. The government assured the first governor of the Japan Development Bank that he would not have to bend to political pressure to fund nonviable projects and that loan decisions would be left to the professional judgment of bank staff and officials. The government was responsible for annually establishing the basic policy for the operation of funds and for conducting regular annual audits. The bank allocated project funds, monitored their spending, and assessed their impact. This system of multiple checks prevented the misuse of policy-based funds and enabled the development bank to keep its loan losses low.

The Japan Development Bank made relatively few mistakes in selecting loan projects. Despite its specialization in long-term industrial finance, the bank's loan losses during the high growth era were much lower than those of the commercial and trust banks, which concentrated on short-term loans and more diversified loan portfolios. The bank experienced write-offs of 0.09 percent of average loans outstanding from 1951 to 1955 and of only 0.01 percent from 1956 to 1965. One caveat should be noted, however. Some of the loan losses incurred in declining industrial sectors, such as coal mining and, later, shipbuilding, were transferred to the Japanese government and absorbed by the general budget. This may explain the unusually low level of loan losses for the period, not only by the development bank, but also by most commercial banks.

When the development bank was established, the equity capital contribution from the government accounted for a substantial portion of its total funds. Because statutory reserves increased in proportion to increases in its loan balance, the bank's financial composition continued to be favorable. It could therefore offer a preferential interest rate in line with policy demands, without being subsidized by public finance. This strong financial position further guaranteed its managerial autonomy.

Empirical Evaluation

Two recent empirical studies lend support to the argument that policy-based finance in Japan was effective in stimulating initial growth and encouraging private investment in growing firms in priority industries. Horiuchi and Sui (1993) compared the investment behavior of medium-size firms receiving development bank assistance for the 1964–88 period with firms of similar size not receiving such funding. They found that the year of initial lending was associated with increased private investment and that within three years, firms began to move away from dependence on development bank lending to rely more on private banks. Horiuchi and Sui also found that directed credit was most effective for firms that did not have main bank affiliations.

Calomiris and Himmelberg (1995) examined the effect of policy-based finance for 1963–91 for the machine tool industry, an industry selected for its high potential for spillover effects through technological innovation and learning. The authors based their study on data for firms, collected with the support of the Japan Development Bank. Because the data set excludes firms that closed during the period, identifying positive effects from policy-based finance is somewhat difficult, particularly for the 1960s and 1970s, when machine tool producers underwent considerable consolidation and less-productive firms shut down or were merged with other firms.

The level of credit sought by machine tool producers declined over the consolidation period. It averaged 27 percent of capital from 1965 to 1974 but fell to 10 percent from 1975 to 1991. Similarly, total long-term debt for these producers fell in relation to capital, from 41 percent before 1975 to 26 percent afterward. Directed credit, which accounted for only a small part of total long-term credit for this period, fell from an average of 3 percent of capital before 1978 to 1 percent after the mid-1980s, or from more than 7 percent of all long term credit to less than 4 percent.

A comparison of lending to general machinery producers by the Japan Development Bank, by private long-term lenders, and by the government-affiliated Export Import Bank of Japan provides interesting insights. Development bank lending declined, from between 3.7 and 5.3 percent of capital in the late 1960s to between 0.8 and 2.6 percent in the 1980s. This clear drop in support is consistent with the premise that government credit relaxed borrowing constraints and helped firms to become seasoned credit risks. Credit from the private Industrial Bank of Japan ranged from between 5.8 and 8.9 percent of capital in the late 1960s to between 4.0 and 6.6 percent in the 1980s. Thus, despite the growing recourse of Japanese firms to the Euromarkets in the 1980s, their reliance on private Japanese funding did not experience the same decline as their use of credit from the development bank.

In contrast to borrowing from the development bank, borrowing from the export-import bank registered a large increase, from between 0.4 and 0.9 percent of capital in the 1960s to between 3.8 and 19.9 percent in the 1980s. The

growing reliance on these loans must be associated with the increasing maturity of the industry and the greater part played by exports, and perhaps overseas operations, in the 1980s. Support from government sources thus showed considerable flexibility in responding to changes in the structure and orientation of the industries and firms receiving credit support. Similar patterns are observed for other types of machine tool producers.

Calomiris and Himmelberg also found that government funds had not been captured at either the industry level or firm level. Directed credit was usually provided to a firm only once and for a brief period (80 percent of firms received one-time credit for an average period of less than eight years). They also found that government credit was provided to large, growing, capital-intensive firms with high investment rates. Directed credit appears to have bolstered the positive characteristics of these firms, reinforcing the process of consolidation, investment, and technological change desired by the government. The authors also found that government credit was positively correlated not only with private credit, but also with reinvestment by the firms themselves. These results are weaker than those Calomiris and Himmelberg reported in 1994 for a shorter period (1982 to 1991). The authors' use of a different methodology in conducting the empirical tests seems to be the main reason for the weaker results reported in 1995. Another reason may be a selection bias in the sample used for the 1963–91 period, which did not include firms that closed in the 1960s and 1970s. If those firms were low investment, poor performance firms and those that survived were the ones more likely to have received government credit, the effect of government credit would be underestimated. In addition, accounting data for the earlier years are probably less reliable.

For the nearly thirty-year sample period studied by Calomiris and Himmelberg, a government loan of 100 yen produced an average investment of 60 yen. The effect of credit from the Japan Development Bank was even stronger: a 100 yen loan produced a 150 yen investment as well as 44 yen in long-term loans from private sources. During the 1960s credit from several government agencies shared equal responsibility for the overall effect of government credit; during the 1970s and 1980s, credit from the development bank had the largest and most significant impact. A possible explanation for this difference may be that other government lenders moved away from producers with high growth potential in the 1970s and 1980s.

Credit Policy in Korea

Credit policy in Korea has also experienced three distinct phases. In the 1950s and 1960s it was directed toward particular activities, mainly exports and industrial investment. In the 1970s it increasingly promoted specific industries, in particular, the heavy and chemical industries. The successes and failures of this phase induced a change in strategy, and in the 1980s the government became

involved in the industrial and financial restructuring of sectors and companies in distress. As in Japan, Korean credit (and industrial) policy began to focus not only on picking winners, but also on phasing out losers. At the same time, policy was reoriented toward producing a more balanced industrial sector that would not be dominated by a few business conglomerates. Lending to small and medium-sized firms received greater attention than it had in the earlier stages, and credit policy was redirected toward functional activities such as research and development and investment in equipment.

Basic Features

Credit policy in Korea differed from that in Japan in several important respects. It involved significant government subsidization of the cost of borrowing, and the scope of directed credit programs was much broader. The Korean program made extensive use of commercial as well as development banks to channel loans to priority sectors; the government owned both types of banks. Government intervention in Korean credit policy was also somewhat coercive. The government used a strong package of tax and financial incentives to encourage firms with minimum equity funds to enter priority industries, and it used its control of the banking system to exert strong leverage on the behavior of firms (World Bank 1987; Cho 1989; Vittas and Wang 1991; Cho and Kim 1995).

In the 1950s credit policies were often implemented without clear industrial policy goals. In the 1960s, however, they were structured specifically to support exports and were linked more closely with other policy measures.

The Korean government undertook a series of measures in the early 1960s to strengthen state control over finance. It nationalized the commercial banks and amended the central bank act to subordinate the Bank of Korea to the government. In 1965 it implemented interest rate reform, doubling the level of bank interest rates. This reform not only prompted the rapid growth of deposits in the government-controlled banks, but also expanded the scope of government control over the allocation of financial resources, as funds shifted from the informal, unregulated market to the regulated sector.

In addition, the government initiated close consultations with the business sector during the 1960s and careful monitoring of the performance of supported firms. The Monthly Export Promotion Meetings and Monthly Briefings on Economic Trends, which were chaired by the Korean president and included senior government officials, managers of industrial firms, bankers, and representatives of industry associations, constituted a forum both among ministries and between the government and the private sector. In these meetings, progress toward achieving policy goals was closely reviewed and a consensus sought on ways to address emerging problems. The economic management of Korean industry thus came to resemble that of a major corporation. The banks were used as the treasury unit; the industrial sector,

as the production and marketing units; and the government, as the central planning and control unit.

During the 1970s the government relied heavily on its control of the credit system, particularly policy-based loans to provide the heavy and chemical industries with preferential access to credit at substantially subsidized rates. The authorities had intended to reduce policy interventions in the credit market in the 1980s, but overexpansion of the heavy and chemical industries in the late 1970s, coupled with the collapse of foreign markets in construction, shipping, and shipbuilding in the early 1980s, forced the government to assist in restructuring industrial firms that faced financial difficulties. As Korean politics became more democratized in the late 1980s, the emphasis of directed credit programs was shifted to social programs and income redistribution.

SOURCES OF FUNDS. A significant difference in the selective credit policies of Korea and Japan has been in the source of funds used for policy loans. Korea has depended heavily on central bank credit and on deposits mobilized by commercial banks, and much less than Japan has on fiscal funds or funds mobilized through the government, such as postal savings. In Korea only 7 to 8 percent of the total value of policy loans extended by commercial banks were financed by fiscal funds; about 35 percent were financed through central bank credit at a discounted rate. Korea's heavy reliance on money creation to finance policy-based loans may in large part explain why prices in Korea have been less stable than in Japan.

Unlike Japan, Korea used foreign capital as a major source of policy-based finance. Foreign capital had a particularly strong effect on Korean economic development, moreover, because domestic savings were far below desired investment levels. Many observers overlook the role of foreign capital in shaping Korean economic policies (including financial sector policies) and the course of development in Korea. Between 1962 and 1982, the average annual economic growth rate was 8.2 percent. A rough estimate suggests that if investment had depended entirely on domestic savings during this period, the average growth rate would have been only 4.9 percent. Without ready access to foreign capital, Korea could not have continued the repressive financial policies that appear to have accelerated economic growth—although these policies also limited the mobilization of financial resources.

The Korean government controlled foreign loan allocation as tightly as it regulated domestic credit. All foreign loans required government authorization, and their allocation was determined by industrial policy goals. In 1965 the government revised the Foreign Capital Inducement Act to allow government-controlled banks to guarantee the repayment of foreign borrowing by firms. These guarantees encouraged inflows of foreign capital and technology, but they also perpetuated government intervention in the banks. To avoid default on foreign loans and possible profound disruptions in development projects, the banks often had to absorb, through the rescheduling of domestic bank loans, the exter-

nal shocks that prevented domestic firms from meeting their foreign debt service. The cost of government intervention in domestic banks had, in turn, to be shared by the banks' depositors.

THE SHARE AND STRUCTURE OF POLICY LOANS. Policy loans mobilized by the Korean financial system were substantial, accounting for about 50 percent of the total credit extended by domestic financial institutions in the 1970s. That proportion gradually decreased to about 30 percent as private nonbank financial institutions, which were not required to make policy-based loans, expanded their share in the 1980s. Policy loans accounted for about 60 percent of total lending made by government-controlled deposit money banks throughout the period.

In the 1960s and 1970s, policy loans were extended mainly to the manufacturing sector; its share of total bank credit was more than twice its share of gross domestic product (GDP), whereas the service sector's share was only about 60 percent of its share in GDP.[1] Within the manufacturing sector, export industries and the heavy and chemical industries received more credit in relation to their share in GDP than did light industries or industries producing for domestic consumption.

Policy Effectiveness

Available data suggest that Korean credit policies were effective in reducing the cost of funds to priority sectors and in enhancing their access to funds. Export-oriented firms enjoyed greater access to credit and lower borrowing costs than did domestic-oriented firms. Heavy and chemical industries, despite their high risk of failure, had greater access to credit and significantly lower borrowing costs than did light manufacturing industries. The availability and low cost of funds aided in the rapid expansion of the export and heavy and chemical industry sectors, especially in the initial, take-off stage.

This evidence does not necessarily imply, however, that selective credit supports were essential for rapid economic growth. The opportunity cost of such supports, that is, the cost of benefits lost by choices not made, are very difficult to estimate. A general equilibrium analysis might provide some answers, but in Korea such an analysis would face severe limitations because substantial parts of input and output prices were controlled in the early stages of development. It is too early, in any case, to provide a full answer to this question. Korean economic development is still in progress, and the cost of past financial policies may not have been fully realized.

It seems clear that export growth drove economic growth in Korea in the 1960s and 1970s and that credit support was indispensable for export growth. Thus credit support must have contributed to rapid economic growth (although a large subsidy may not have been necessary to trigger export growth). The effect of credit supports on growth of the heavy and chemical industries remains controversial, however. Although credit supports contributed to the rapid de-

velopment of the two industries, credit might have been used more efficiently, given the labor endowment in the 1970s, if its allocation had been better balanced between the heavy and chemical industries and light industry. It can also be argued, however, that had the push to develop the heavy and chemical industries not taken place in the 1970s, and the two sectors not become Korea's leading export industries in the mid-1980s, Korea might not have been able to take full advantage of the appreciation of the Japanese yen and the world economic boom in the second half of the 1980s. No solid answer can be provided to this question.

The impact of credit policies on industrialization is not limited to their effect on the cost of and access to credit. In an economy like Korea's, in which initial capital accumulation was low, and rapid investment growth was financed mainly by bank credits and foreign loans, firms were highly leveraged financially. As a result, any significant economic downturn would precipitate a financial crisis unless schemes existed for sharing risk between creditors and borrowers. By controlling finance, the Korean government became an effective risk-sharing partner with the industrialists. It could thus encourage their venturesomeness and induce them to have long term horizons for their companies. The government was thus a partner, with industry and the banks, in an implicit coinsurance scheme. Without this partnership, Korea might not have been able to establish its large, internationally competitive industrial firms so quickly. This indirect effect of the government's credit policies may be more important than the subsidies themselves in explaining Korea's rapid industrialization.

The Cost and Legacy of Credit Policies

Government control of finance in Korea was not without cost. Several commentators have argued that the drive to develop the heavy and chemical industries was overambitious and resulted in a serious misallocation of resources. These critics maintain that the priority sectors expanded production capacity too rapidly, giving firms too little time to accumulate experience and digest new technologies. The result, critics say, was that most firms experienced excess capacity, high production costs, and products so low in quality that they could not be exported. At the same time, nonpriority sectors were forced to borrow at very high rates from the informal sector. This dual nature of the credit system created a major imbalance in the industrial structure of the country (Koo 1984; Kwack 1984).

Because Korea relied too heavily and for too long on credit interventions as an instrument of industrial policy, the costs were borne primarily by banks and their depositors. Commercial banks in Korea were involved so deeply in directed credit programs that they functioned as development banks. Managerial efficiency and quality of services were compromised, and the banks had large volumes of nonperforming loans. Nonbank financial institutions, which operated more freely, expanded rapidly and superseded the banks' share in the fi-

nancial intermediation market. The expansion of the nonbanks helped to improve financial market operations by keeping competitive forces alive in the financial system.

The problem of carelessness and irresponsibility at the commercial banks was no less serious. As long as the government was willing to rescue firms, banks had little incentive to screen projects carefully or to monitor firms' performance. Indeed, the firms supported by the government became too large and dominant to be allowed to go bankrupt. In the mid-1980s when export markets collapsed after the second oil shock and the worldwide recession, considerable financial support was given to ailing firms and industrial sectors. Although detailed data for this period are not readily available, these setbacks apparently made it increasingly difficult for the government to break out of a pattern of financial repression. When the subsequent expansion of the Korean economy and the increasing sophistication of its industrial structure called for a more innovative and market-oriented financial sector, this past legacy became a constraint on liberalization.

Because Korea's industrial policy emphasized the economies of large-scale production to maintain international competitiveness, it led to overwhelming economic concentration within huge business conglomerates, known as *chaebols*. During the 1970s, it was not uncommon for these conglomerates to triple the number of their affiliates through new acquisitions. In the 1980s, in the face of growing public discontent with excessive economic concentration, the government was forced to redirect policies toward redistributing income; this change in focus often involved increased regulation of the business activities of the large firms.

Factors of Success in Japan and Korea

Theory and practical experience in Japan and Korea suggest that credit policy can be an effective policy instrument for economic development. Why, then, did credit policies fail in so many countries, and what factors explain their relative success in Japan and Korea?

The presence of specific economic and institutional factors seem to be essential to success. Economic factors include the maintenance of price stability, an orientation toward export production, the encouragement of domestic competition, a reliance on the private sector, and a bias toward industrialization. Institutional factors include the use of extensive public-private consultative arrangements, the creation of effective monitoring systems, and, above all, the development and propagation of credible visions.

Price Stability

Japan and Korea have, with occasional exceptions, both been able to maintain price stability. Price stability in itself, however, does not seem sufficient to

ensure the success of credit policies. Several countries in the Middle East, North Africa, South Asia, and southern Europe have also avoided high inflation rates. India, for example, has maintained better price stability than has Korea, but neither India's credit policies nor its economic performance have been as successful as Japan's or Korea's. Price stability seems clearly to be important for encouraging the growth of financial savings. India and other countries with moderate inflation have experienced a substantial deepening of their financial sectors compared with countries with similar levels of per capita income but with higher inflation. Japan, with very stable prices, has experienced steady and rapid growth of its financial system. Korea, however, with relatively high inflation, experienced poor growth of its financial sector in the 1970s, although that trend was reversed in the 1980s. The recent rapid growth of financial saving in China, where the macroeconomic environment has been relatively stable compared with other developing countries, also supports the importance of price stability.

Competition, Exports, and Private Ownership

The credit policies of Japan and Korea have been more successful than other countries primarily because those two countries were able to combine price stability with intense domestic competition, strong export orientation, and a significant reliance on the private sector. A strong orientation toward exports has forced domestic firms in Japan and Korea to attain high levels of efficiency in order to compete internationally. It has also provided objective criteria for monitoring the performance of individual firms and for assessing the effectiveness of credit supports. Good performance in export markets, for example, implies continuing access to policy-based finance.

Although both Japan and Korea have had import protection, the large industrial firms have engaged in vigorous competition in their respective domestic markets. Even in the economically stable countries of the Middle East, South Asia, and southern Europe, industrial production has often been sheltered from both domestic and foreign competition, has been directed toward domestic consumption, and has been controlled by state-owned enterprises.

Industrialization and Export Promotion

Credit policies in Japan and Korea have been narrowly focused and well coordinated with other policies. Policy measures, such as those dealing with foreign exchange, tax, and fiscal arrangements, have been directed toward supporting industrialization and export promotion, the main objective of credit policies. The allocation of foreign exchange in Japan and the approval of foreign loans in Korea have been coordinated with domestic credit policies to advance industrial strategy. Credit policies in India, by contrast, have focused on redistributing wealth and income toward small farmers and firms and have thus oft-

set other programs supporting manufacturing and exports. These conflicting policies have often resulted in an implicit taxation of large industries to support farmers and small business.

Consultation with the Private Sector

Government participation and leadership in organizations such as the deliberative councils in Japan and the industrial associations in Korea have been critical in helping these countries avoid the pitfalls of credit policy. The cooperative networks they have established have helped to ensure that scarce resources are allocated to activities likely to yield long-term benefits across the economy, and they have provided an effective mechanism for collective risk sharing by industry, government, and their lenders. Monthly meetings have offered opportunities to collect and exchange information, to reassure individual firms and lenders of the government's commitment to particular enterprises, and to promote consensus in pursuit of industrial policy goals. Organizations such as these, however, may be subject to manipulation, corruption, and inefficiency. Their use in allocating credit is an improvement over market solutions only if there are strong safeguards against abuse.

Monitoring

Reference has already been made to the effectiveness of monitoring in Japan and Korea. Loan approval in Japan is preceded by careful review and independent appraisal, and the use of funds is strictly monitored. Fund disbursement is based on adequate documentation, and continued access to policy-based loans depends on success in reaching objective targets, mostly in internationally competitive export markets.

The importance of effective monitoring cannot be overemphasized. What matters for economic development is not really the mobilization and allocation of financial resources but the efficient use of those resources. Any country can mobilize resources by printing money, which it can then allocate to priority sectors. Such money creation will, of course, increase inflationary pressures, but if the resources are used well and lead to higher and more efficient levels of production, unit costs will be lowered and will offset the inflationary impact of the initial monetary financing.

Vision

Although economists have generally paid little attention to the broad aspects of strategy, the importance of having a credible and consistent vision for economic development cannot be overstated. In both Japan and Korea, industrialization and economic growth have taken precedence over the development of an efficient and modern financial sector. Although the authorities have been com-

mitted to ensuring the safety of deposits and the solvency of financial intermediaries, they have been less concerned with allowing banks and other financial intermediaries to innovate and develop new services intended to reduce the cost of financial intermediation. Banks in both countries have been subject to extensive controls—on bank spreads and interest rates, branching and bank mergers, and bond issues—as well as to administrative guidance that initially discouraged lending for consumer credit and housing finance and encouraged the creation of large industrial-financial groupings. In addition, both countries have encouraged lending to fast-growing, high-yield industries, such as automobiles, electronics, petrochemicals, and steel, in which income elasticity of demand has been high, technological progress has been rapid, and labor productivity has risen quickly. Their strategies have emphasized dynamic comparative advantage rather than static cost considerations (Ojimi 1972; Johnson 1982; Yotopoulos 1991).

It is ironic that the current sorry state of Japanese banks derives from bad loans made after credit policies were relaxed in the 1980s and real estate lending was expanded exponentially. The lesson to draw from the current experience of Japanese banks, however, is not that liberalization was ill advised. Japan's industrial success made inevitable a reorientation of its financial system toward real estate lending, housing finance, and consumer credit. What hurt the banks was the failure to recognize accumulating problem loans and to take early measures to address the impact of that buildup on the banking system.

Several other aspects of broad strategy have contributed to the success of credit policies in Japan and Korea. Both countries, for example, have emphasized *complementarities in production* for exports and domestic sales. This tactic has allowed resources to shift to exports when balance of payments problems forced the government to curtail domestic demand and to shift back to domestic sales when the problems of paying for imported raw materials have eased. The rationale is that if factories can operate through all phases of the business cycle, they can achieve higher scales of production and lower operating costs (Johnson 1982, Yotopoulos 1991).

Japan has benefited from an industrial strategy that has been adjusted over time to meet the needs of general development goals. This *flexibility of focus* has allowed Japan to emphasize, first, the recovery of priority production, then the modernization of equipment in heavy industries, and, later, the development of new industrial sectors, such as machine tools, with a high potential for producing positive economic side effects. The strategy has also accommodated the smooth adjustment of declining industries, the restructuring of companies facing difficulties, and the rationalization of entire sectors of industry suffering from overcapacity.

Credit policies in Japan and Korea have also emphasized the use of *additional instruments of industrial policy* to complement policy-based finance. Both countries have made extensive use of incentives such as accelerated depreciation allowances and tax-free special reserves. These incentives have permitted success-

ful firms in targeted sectors to retain and reinvest a larger part of their profits than firms in nontargeted sectors have been allowed to retain. Particularly important have been the special reserves that are tied to past export performance, because these are linked with the overall strategy of export promotion. These incentives have reinforced the impact of credit policies and have accentuated the credibility of directed credit programs.

The existence of coherent, credible visions for development in Japan and Korea has also lent substance *to the consultative process*. Many countries have tried to promote close consultation between government and the private sector, but without a clear vision for development, such exchanges have been little more than ineffective talk shops or forums for special pleading. An important contribution to this collaboration in Japan and Korea was the governments' collection and dissemination of information about long-term sectoral prospects, an activity not readily undertaken by the private industrial sector or the private securities markets. A clear vision of economic goals supported by the analysis of broadly reliable data about the prospects of particular industrial sectors reinforced the economic signaling effect of policy-based finance in Japan and Korea.

Lessons from Japan and Korea

Ten general lessons can be drawn from the experiences of Japan and Korea. The first six relate to "good vision," the last four to "good management." First, credit programs must be small in size, have a narrow focus, and be of limited duration with clear phase-out provisions. Second, programs must involve a low level of subsidy (if any), to minimize incentives to profit from price fluctuations and also to reduce the tax on financial intermediation that all credit programs necessarily entail. Third, programs must be financed by long-term funds to avoid inflation. Recourse to central bank credit, in particular, should be avoided except in the very early stage of development, when selective credit programs supported by central bank credit might help jump-start economic growth and development. Even at this stage, however, care must be taken to prevent high rates of inflation. Fourth, programs should aim at achieving positive economic or social side effects (or at avoiding negative effects). They should therefore focus on overcoming the external finance constraint facing small or rapidly expanding firms as well as helping firms in declining industries to phase out in an orderly and timely fashion. Fifth, programs should promote industrialization and export orientation, and they should be based on a competitive private sector with internationally competitive operations. Sixth, programs should form part of a broader credible vision of economic development, promoting growth with equity and including a long-term strategy to develop a sound financial system operating on economic criteria.

Seventh, policy-based loans should be channeled through well-capitalized, administratively capable, and autonomous financial institutions. Professional

management and managerial autonomy are essential. Eighth, loans should be based on clear, objective criteria that are easily monitored. Detailed project appraisals, close supervision of disbursements, and monitoring of performance and repayment records are essential to the success of credit programs. Ninth, policy-based programs should aim for a good repayment record and low loan losses. Tenth, programs should be supported by effective mechanisms for consultation between the public and private sectors, including the collection and dissemination of basic market information.

Although these lessons are important, replicating the Japanese and Korean experience may be difficult in today's financial environment. High technology coupled with the globalization of financial markets has substantially reduced the effectiveness of foreign exchange controls on capital movements and has limited the ability of government authorities to set interest rates substantially below market levels. The new World Trade Organization agreement, moreover, limits the use of credit policies. Despite these potential obstacles, however, the greatest challenge facing most developing countries in the use of policy-based finance remains the absence of the essential factors of good vision and good management that explain the success of directed credit policies in Japan and Korea.

Notes

Dimitri Vittas is Adviser with the Financial Sector Development Department of the World Bank. Yoon Je Cho is Vice President of the Korea Tax Institute. This article draws on the findings of a World Bank research project on the "Effectiveness of Credit Policies in East Asian Countries." These findings were published either as World Bank Discussion Papers or as Policy Research Working Papers and are cited in the list of references. We are grateful to our many collaborators in this project and to several staff members of the World Bank for useful insights and comments.

1. The manufacturing sector received 46 percent of total bank loans given in 1970 and 54 percent of the total given in 1980. In contrast, the service sector received only 29 percent in 1970 and 24 percent in 1980.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Calomiris, Charles W., and Charles P. Himmelberg. 1994. "Directed Credit Programs for Agriculture and Industry: Arguments from Theory and Fact." *Proceedings of the World Bank Annual Conference on Development Economics 1993.* Washington, D.C.: World Bank.

———. 1995. "Government Credit Policy and Industrial Performance: Japanese Machine Tool Producers, 1963–1991." Policy Research Working Paper 1434. World Bank, Financial Sector Development Department, Washington, D.C.

Calomiris, Charles W., Charles P. Himmelberg, Charles M. Kahn, and Dimitri Vittas. 1992. "Evaluating Industrial Credit Programs in Japan: A Research Proposal." Working Paper, University of Illinois at Urbana-Champaign, Department of Finance. Processed.

Cho, Yoon Je. 1989. "Finance and Development—The Korean Approach." *Oxford Review of Economic Policy* 5(4, Winter):88–102.

Cho, Yoon Je, and Thomas Hellmann. 1994. "The Government's Role in Japanese and Korean Credit Markets—A New Institutional Economics Perspective." *Seoul Journal of Economics* 7(4, Winter):383–416.

Cho, Yoon Je, and Joon-Kyung Kim. 1995. *Credit Policies and the Industrialization of Korea.* World Bank Discussion Paper 286. Washington, D.C.

Horiuchi, Akiyoshi. 1984. "The 'Low Interest Rate Policy' and Economic Growth in Post War Japan." *The Developing Economies* 22(4, December):349–71.

Horiuchi, Akiyoshi, and Qing-Yuan Sui. 1993. "Influence of the Japan Development Bank Loans on Corporate Investment Behavior." Discussion Paper Series 93-F-4. University of Tokyo, Faculty of Economics. Processed.

Ikeo, Kazuhito. 1987. "Japan's Financial System: A Micro Approach." *Japanese Economic Studies* (Fall):60–77.

JDB/JERI (Japan Development Bank/Japan Economic Research Institute). 1994. *Policy Based Finance: The Experience of Postwar Japan* World Bank Discussion Paper 221 Washington, D.C.

Johnson, Chalmers. 1982. *MITI and the Japanese Miracle.* Stanford, Calif Stanford University Press.

Koo, Bohn-Young. 1984. "Role of the Government in Korea's Industrial Development" Working Paper 84–07. Korea Development Institute, Seoul October

Kuroda, Iwao, and Yoshiharu Oritani. 1980. "A Reexamination of the Unique Features of Japan's Corporate Financial Structure—A Comparison of Corporate Balance Sheets in Japan and the United States." *Japanese Economic Studies* (4):82–117

Kwack, Taewon. 1984. "Industrial Restructuring Experiences and Policies in Korea in the 1970s." Working Paper No. 84 08. Korea Development Institute, Seoul

OECF (Overseas Economic Cooperation Fund) 1991 *Issues Related to the World Bank Approach to Structural Adjustment—Proposal from a Major Partner* Occasional Paper 1 Tokyo.

Ojimi, Yoshihisa. 1972. "Basic Philosophy and Objectives of Japanese Industrial Policy In Organization for Economic Cooperation and Development, *The Industrial Policy of Japan*. Paris.

Suzuki, Yoshio. 1980. *Money and Banking in Contemporary Japan.* New Haven, Conn Yale University Press.

Vittas, Dimitri, and Akihiko Kawaura. 1995. "Policy-Based Finance, Financial Regulation and Financial Sector Development in Japan" Policy Research Working Paper 1443 World Bank, Financial Sector Development Department, Washington, D C

Vittas, Dimitri, and Bo Wang. 1991. "Credit Policies in Japan and Korea—A Review of the Literature." Working Paper Series 747. World Bank, Country Economics Department, Washington, D.C.

World Bank. 1987. *Korea: Managing the Industrial Transition* Washington, D C

Yotopoulos, Pan A. 1991. "Exchange Rates and State Led Capitalism What Can the NICs Learn from Japan?" *Asian Economic Journal* 5(1):76–99

PRODUCER TURNOVER AND PRODUCTIVITY GROWTH IN DEVELOPING COUNTRIES

Mark J. Roberts
James R. Tybout

The reallocation of resources, either across sectors or across producers within a sector, can serve as a potential source of productivity growth. New research findings exploit comprehensive microeconomic data on the manufacturing sectors of Chile, Colombia, and Morocco to document resource shifts as producers enter, expand, contract, and exit operation. The micro-level adjustment is substantial; between 25 and 30 percent of the total number of manufacturing jobs turn over each year. In the short run, the productivity effects of this turnover are modest because the new plants that come on line are only slightly more productive than the ones they replace—and both are typically small. In the longer term, however, the turnover generates more substantial increases in productivity because the new firms that survive record substantial productivity gains in their early years. Moreover, firms that exit are typically on a downward productivity spiral and would probably have dragged down sectoral efficiency farther if they had continued in operation.

Many developing countries in the process of structural transformation are struggling to catch up technologically, and their labor markets face the formidable task of moving workers among diverse occupations. At the same time, various frictions inhibit factor mobility in the industrial sector; these include severance laws that prevent firms from firing workers and regulations that limit the establishment of new firms and the termination of old ones. The movement of capital and labor is further constrained by credit-market imperfections, noncompetitive markets, and limited information about technological advances. Many resources might earn higher social returns if they were redeployed in different activities, and structural changes that work to this end might generate large efficiency gains. Similarly, if rates of productivity growth differ across sectors, resource shifts

The World Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 1–18

from low- to high-productivity sectors might significantly improve productivity growth throughout the economy.

This notion that growth can be generated by reallocating resources more efficiently has long interested development economists. It is well documented that broad shifts in the composition of output take place as the development process unfolds, typically shifting production away from natural resource-based products and toward manufacturing and service sectors (Chenery 1979; Chenery, Robinson, and Syrquin 1986; and Syrquin 1988). Several studies have found that this structural change is often associated with substantial gains in productivity, but little is known about the resource reallocations that occur *within* sectors as producers enter and exit and their respective market shares change. As Kuznets (1979) and Syrquin (1984) note, if these processes move resources from less efficient to more efficient plants within the same sector, the gains in productivity may be substantial, but the source of these gains is impossible to identify using aggregate data.[1]

This paper summarizes recent research on the magnitude and implications of this micro-level reallocation of resources in semi-industrial countries. A common theme of this research is that individual producers within the same industry differ in efficiency and thus the reallocation of resources from less efficient to more efficient producers offers the potential for improved economic performance. The research is based on panel data that cover virtually all the manufacturers with at least 10 workers in three countries: Chile, Colombia, and Morocco. (Other, more limited data from Mexico, Turkey, and Venezuela are also used.) These data make it possible to track individual producers as they enter, expand or contract production, and exit. Differences in productivity can be measured, and the effect of micro-level resource reallocations on aggregate productivity can be quantified.

Several robust patterns of resource reallocation exist. First, a tremendous amount of adjustment at the micro level is completely masked in aggregate sectoral data. Each year the entry of new manufacturing plants and the growth of existing plants create new jobs that average between 13 and 19 percent of total employment in the manufacturing sector. At the same time the contraction and closing of other plants are responsible for the simultaneous loss of between 12 and 14 percent of total employment. This high rate of employment reallocation among manufacturing plants is present in all the countries studied and persists in each year throughout the business cycle, reflecting a vigorous process of micro-level adjustment.

Second, the patterns of plant turnover partly reflect differences in productivity across plants. On average, exiting plants are less productive than surviving ones, and entering plants are less productive than more experienced incumbents. As new plants mature, however, their average productivity tends to increase for several years until they reach industry norms. Overall, the empirical results reveal a continual process of resource reallocation that moves resources from less-efficient to more-efficient producers within the same industry and that contributes to long-run improvements in economic performance.

In this article we first review the empirical evidence on patterns of producer mover, describing the creation and destruction of jobs and the entry, growth, d exit patterns of manufacturing plants. We then summarize the evidence on ferences in productivity among plants and the implications of turnover for oductivity growth. Finally, we address some of the reasons for the differences efficiency across producers.

oducer Turnover and Employment Flows

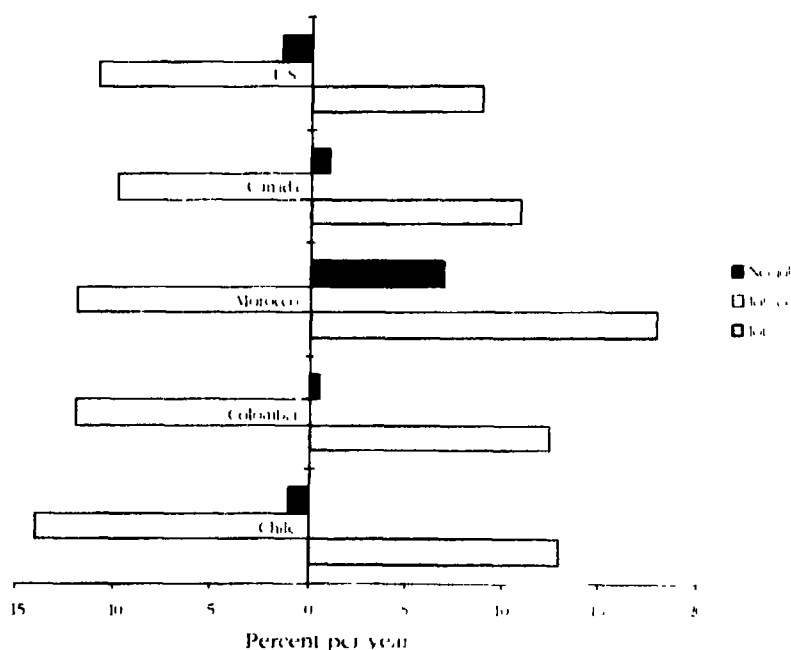
It is useful to think of the reallocation of resources as arising from three fferent forces. The first source, which has received the most attention in the erature on development, is long-run structural shifts in technology, endownts, and demand. These forces generate an expansion of output and attract w producers in some sectors, while inducing a contraction in output and net it by producers in other sectors. Typically, industrial sectors in developing untries gradually shift out of assembling low-technology manufactured goods d move into more sophisticated products that are relatively intensive in huar and physical capital.

The second source of resource reallocations is short-run or cyclical fluctuans in demand that might arise from changing macroeconomic conditions or ternal market shocks. These cyclical fluctuations may be felt in all sectors, but eir effects are likely to be particularly important in industries that have low ck costs because in such cases short term, or hit-and-run, entry may be profble. In industrial countries cyclical patterns of job creation and destruction ve been interpreted to have a "cleansing" effect because in recessions resources released from those activities that produce the lowest return and are subseently reemployed more productively when the economy expands (Caballero d Hammour 1994).

The third source of resource reallocation stems from market forces that crecontinual producer turnover within an industry even when macroeconomic ditions are stable. This phenomenon has only recently been formally moded in the literature (see Jovanovic 1982; Lambson 1991, 1992; Hopenhayn 99 and Pakes and Ericson 1995). It derives from the fact that sunk entry and it costs, combined with uncertainty, make it possible for producers at differt levels of efficiency to co-exist in the same industry. (Entry costs include censing fees and irreversible purchases of capital goods; exit costs may include nkruptcy expenses or severance payments to employees.) Differences in efficiency arise from differences in managerial abilities or random variation in the turns on past investments in capital or technology. A given firm is uncertain bout its relative efficiency and learns about it through market experience or by bserving the outcomes of investment projects. Those companies that find they re relatively inefficient or that invest in unproductive assets eventually exit, hile those that find they are efficient or that their investments are productive

survive and expand. New firms continually enter and try their hand at com ing with the incumbents. The speed of this turnover process is affected by r ket conditions and by the magnitude of the sunk entry and exit costs invol Institutional frictions such as severance laws and credit constraints also sł turnover patterns.

To study all three types of resource reallocation requires examining all producers in a sector—not simply large continuing producers. Comprehen panel data from Chile, Colombia, and Morocco were used to identify en ing, incumbent, and exiting producers, impute productivity trajectories each producer, and calculate market shares. Figure 1 summarizes the ma tude of annual job creation and destruction in the manufacturing secto each country.[2]

Figure 1. *Job Creation and Destruction*

Note Sample periods are Chile 1979–86 Colombia 1977–89 Morocco 1984–89 [illegible] and United States 1973–86. The annual rate of job creation is the number of employee positions added during a year, expressed as a percentage of total manufacturing employment at the start of the year. Similarly, the annual job destruction rate is the number [illegible] lost during the year, expressed as a percentage of employment at the beginning of the year. The annual rate of employment growth in the manufacturing sector, referred to as net job creation, is the difference between jobs created and jobs destroyed.

Source. Roberts and Tybout (1996, ch 2, table 2.1)

Economic conditions in these three countries were typical of those in most [illegible]-industrial countries in the 1980s, so the findings are probably representa[illegible] of a broader group of nations. Each of the three countries began the decade [illegible] an overvalued currency and was forced to devalue and contract during the [illegible] crisis. By the end of the sample period, each country had undergone some [illegible]gree of structural adjustment and resumed growing.

There were, however, important differences among the three. Chile suffered [illegible] financial crisis in the early 1980s because the manufacturing sector had [illegible] heavily indebted in dollars. Its contraction was severe, with unemploy[illegible] reaching almost 30 percent and large-scale shutdowns of manufacturing [illegible]. Nonetheless, economic policies remained laissez faire, with low tariffs, [illegible] no nontariff barriers, very little public ownership in manufacturing, and [illegible] intervention in the labor market. Colombia's recession was much milder, [illegible] its commercial policy remained more protectionist. The data base for Mo[illegible] does not begin until after the recession, so it describes only the prolonged [illegible]. During that time the government promoted manufactured exports with [illegible] tax exemptions but maintained some degree of protection from imports. [illegible] differences, as well as variations in the length of the sample periods and [illegible] of industrialization in each country, probably led to some differences [illegible] reallocation of resources over the long term and in the volume of intra-[illegible] turnover.

[illegible] Fluctuations

[illegible] way to quantify micro-level resource flows is to look at the creation and [illegible] of jobs (see figure 1). In Chile, Colombia, and Morocco, new jobs were [illegible] at an annual average rate of 13 to 19 percent of total employment in manu[illegible], while the average rate at which jobs were eliminated varied from 12 to 14 [illegible]. These rates were remarkably similar across the three countries, considering [illegible] different macroeconomic conditions that prevailed. Together, the average [illegible] of new manufacturing positions that were added and existing positions that [illegible] came to 26 to 30 percent of total manufacturing employment in an aver[illegible] somewhat more volatility than one finds in the United States and Canada. [illegible] explanation for this result is that macroeconomic shocks were longer in the [illegible]-industrial countries. Nonetheless, annual figures (table 1) show that most job [illegible] and destruction takes place simultaneously at all phases of the business [illegible], implying that inter- and intra-industry turnover together are more important [illegible] the effects of the business cycle.

[illegible] the costs of changing employers are similar for workers in semi-industrial [illegible] and in the United States and Canada, the rapid rate of job turnover implies that the average adjustment burden per worker is relatively high. These costs may be somewhat offset, however, by the high geographic concentration of manufacturing activity in the semi-industrial countries, which makes it less likely that workers will need to move as employment demand shifts.

Table 1. *Job Creation and Job Destruction by Phase of the Business Cycle*
(percent)

Country	*Rate of job growth*	*Gross job additions*	*Gross job losses*
	Average during years of employment expansion		
Chile	8.7	17.6	-8.9
Colombia	2.8	13.7	-11.0
Morocco	6.5	18.6	-12.1
Canada	2.6	11.4	8.8
United States	3.2	11.1	7.9
	Average during years of employment contraction		
Chile	-8.2	9.4	17.6
Colombia	-2.2	11.2	-13.3
Morocco	n.a.	n.a.	n.a.
Canada	3.0	9.1	12.1
United States	-5.5	7.4	12.9

n.a. = Not available.

Note: Sample periods are Chile, 1979–86; Colombia, 197[illegible]–89; Morocco, 1984–89; Canada and [illegible] States, 1973–86.

Source: Roberts and Tybout (1996, ch. 2).

A comparison of the annual job creation and job destruction rates betw[illegible] the semi-industrial countries and the United States and Canada reveals an in[illegible] esting contrast. As the two industrial countries move from recession to exp[illegible] sion, the reduction in the rate of job destruction is larger than the increase in [illegible] rate of job creation. As a result total employment turnover is countercyclica[illegible] and consistent with the view noted earlier that recessions "cleanse" the prod[illegible] tion structure (Caballero and Hammour 1994). In the semi-industrial countr[illegible] job creation rates are equally—or more—sensitive than job destruction rates [illegible] fluctuations in aggregate economic activity. That is, the job creation rate diff[illegible] more between expansionary and contractionary periods than does the job [illegible] struction rate. Thus, the dominant cyclical feature is the large increase in [illegible] creation of new job opportunities during expansionary periods. One interpre[illegible] tion is that limited access to financial markets in the semi-industrial countr[illegible] forces plants to rely more heavily on internal finance, so expansion and [illegible] are more sensitive to demand.

Structural Shifts

Despite the high turnover during the sample periods shown in figure 1, re[illegible] tively little change occurred in the net size of the manufacturing sector in ea[illegible] of the three semi-industrial countries. Over the entire sample period, total man[illegible] facturing employment changed by an average of only 0.3 percent a year in Co[illegible] lombia, −1 percent in Chile, and 6.5 percent in Morocco. Especially in Chile an[illegible] Colombia, therefore, intersectoral labor reallocation between manufacturing an[illegible]

he rest of the economy (including the pool of unemployed workers) accounted or a small fraction of gross job reallocation flows. This is presumably because he time periods examined are relatively short and the sample countries were urly industrialized at the beginning of the sample period. In any case, the data do not show the kind of broad intersectoral reallocation of jobs noted in earlier tudies.

Yet the industrial sectors of developing countries continually change charac-r as labor-intensive, light-manufacturing industries give way to more capital-tensive, durable goods industries (Chenery, Robinson, and Syrquin 1986). One ight therefore expect a substantial shift in labor flows across industries *within* e manufacturing sector, even though aggregate manufacturing employment is ot changing much. But we found no more shifting of jobs from one manufac-ring industry to another in the semi-industrial countries than in the United States. After controlling for the net expansion or contraction of total manufac-ing employment, we found that more than 80 percent of the shift of workers ployed in manufacturing occurred *within*, rather than across, industries. That s, shift in positions from plants that are contracting or failing to plants that *rering or expanding in the same manufacturing industry* accounts for more 80 percent of the annual change in employment on average. Presuming t worker skills are industry specific rather than employer-specific, this turn-er pattern implies that displaced workers require less retraining than they ould if they moved to another sector altogether.

ra-Industry Turnover

he dominance of internal flows of employment within the same industry ggests that industrial evolution models best describe the data. If this is so, ch of the job creation and destruction reflects the continual exit of producers are relatively inefficient and the continual entrance into the same industry ew producers who are, on average, better. Another implication is that sunk ry and exit costs, which are largely dictated by the capital requirements of uction, are fundamental determinants of the speed at which this cleansing ess unfolds. Producers are reluctant to enter industries with high sunk costs, umbent producers are less likely to be driven from the market, and on age, these industries will purge inefficient plants relatively slowly. Further, se same industries most of the output is controlled by a few producers, and ing profits are high, even when incumbents have no market power ovic 1982). This pattern tends to support the conclusion that conditions tionally associated with monopoly rents may also be consistent with com-tive behavior and that antitrust action may not be warranted on welfare unds.

Several other patterns in the data suggest the importance of sunk entry and exit costs. First, the rate of job turnover, defined as the sum of the rates of job creation and destruction, differs substantially across industries, but the ranking

of industries from low to high turnover tends to be very similar across countr
Industrial technology, which is common to all countries, appears to play a la
role in shaping this pattern. Second, the high-turnover industries, such as fu
ture, apparel, food processing, and wood products, are all ones with relati
small-scale production and low capital intensity, while the low-turnover ind
tries, such as steel, chemicals, glass, and paper, are the opposite (table 2).

Job turnover resulting from entry and exit as well as from expansions and c
tractions in the size of existing plants reflects managerial reactions to the fi
success or failure in the market or the outcomes of investments—or both. The
source reallocations that accompany each process, however, are subject to diffe
types of frictions. For example, because sunk costs are associated with opening
closing a business, the decision to enter is forward-looking and reflects expecta
about the entire future profit stream. Adjustments by incumbents in the numb
workers employed and the volume of materials purchased are driven mainly
current profit considerations, however. Policies such as bankruptcy laws, entr
censing requirements, and severance laws affect both entry and exit decision
decisions on the scale of operations, but in different ways

Although entry and exit account for about a third of total job turnove
Chile and Morocco, and nearly half the job turnover in Colombia, they are
important in the United States (table 3) In fact, when business-cycle effects
netted out, the difference between turnover rates in the United States and th
in the semi-industrial countries can be attributed entirely to the latter gre
rates of entry and exit. These higher rates, in turn, trace to the relative emph
in those countries on light manufacturing industries, in which entry and
costs are small. If low entry costs lead to strong competitive pressures, th

Table 2. *Average Annual Employment Turnover Rates by Three Digit ISIC Industry*

Industry	*Percentage turnover*	*Industry*	[illegible]
Iron and steel	11	Professional/scientific equipment	[illegible]
Industrial chemicals	12	Printing	[illegible]
Glass	12	Nonmetallic mineral products	[illegible]
Ceramic products	12	Leather	[illegible]
Paper	13	Plastic products	[illegible]
Rubber	14	Footwear	[illegible]
Beverages	14	Fabricated metal products	[illegible]
Nonferrous metal refining	14	Nonelectrical machinery	[illegible]
Electrical machinery	16	Furniture	[illegible]
Transport equipment	16	Apparel	[illegible]
Other chemical products	16	Food processing	[illegible]
Textiles	18	Wood products	[illegible]

Note: Definition of industry group based on the International Standard Industrial Classification (ISI
designed to promote international comparability in statistics of economic activity

Source: Roberts and Tybout (1996)

Table 3. *Entry and Exit as a Source of Job Turnover*

(percent)

Country	Total job turnover	Job turnover due to plant entry and exit	Job turnover among incumbent plants
	Average during years of employment expansion		
Chile	26.5	8.9	17.6
Colombia	24.7	11.9	13.0
Morocco	30.8	8.9	21.8
United States	19.0	3.5	15.5
	Average during years of employment contraction		
Chile	27.0	10.4	16.6
Colombia	25.5	11.3	13.2
Morocco	n.a.	n.a.	n.a.
United States	20.3	4.0	16.3

n.a. Not available.

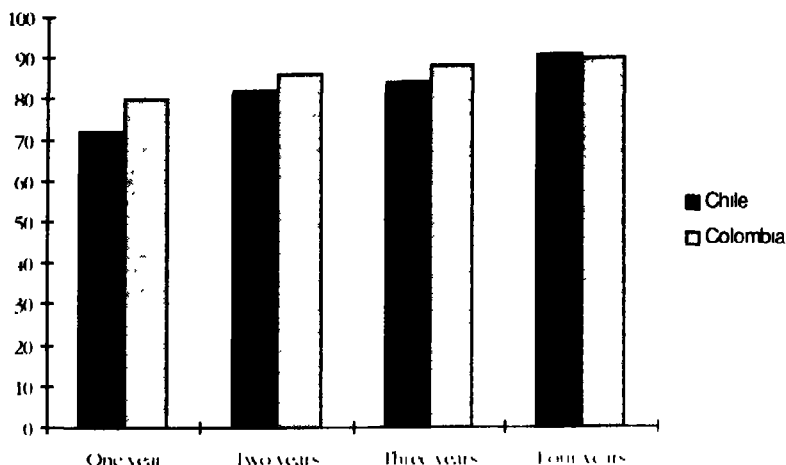
Source: Roberts and Tybout (1996).

...turnover rates challenge the popular perception that the high concentration of manufacturing in many developing countries reflects less market competition than is seen in industrial countries (Rodrik 1988, Krugman 1989).

One view of the life cycle of individual plants is that new entrants embody the latest technology and thus are more efficient than older plants. These new plants gradually expand and become significant producers but are eventually replaced, in turn, by producers embodying newer technology. Under this scenario, exiting plants should be among the oldest because they are likely to be relying on outdated technology. An alternative view, however, suggests a contrasting pattern. If plants are born with different efficiencies and only learn their relative efficiencies gradually as they gain experience, it follows that the efficient ones survive and grow, while the inefficient ones contract and exit. In this case, exit would be concentrated among younger plants, while older plants, having survived the shakedown process, will be the most efficient.

Data on failures by manufacturing plants in semi-industrial countries support the latter view. Figure 2 shows the share of new manufacturing plants in Chile and Colombia that survived each year during their first four years of operation. As each cohort of new plants ages, the proportion of plants that survive steadily increases. For example, the one-year survival rate in Colombia increased from 69.4 percent for one-year old plants to approximately 87 percent for plants more than three years old. In Chile, first-year survival rates averaged 73.2 percent, while four-year-old plants had survival rates of 89.2 percent.

Additionally, the average size of the surviving members of the cohort increases over time. For example, in a typical year, one-year old plants in Chile and Colombia are only 26 percent and 39 percent as large, respectively, as the average incumbent plant, but five-year-old plants are 75 percent and 65 percent as large,

Figure 2. *Survival Rate of Manufacturing Plants*
(percent)

Source: Roberts and Tybout (1996, ch. 7, 8)

respectively. This increase reflects two effects: the growth of the surviving co-hort members and the failure of the smaller plants in the cohort. Both factors, however, indicate that in each year, it is the older plants that are the dominant source of industry output.

Overall, the qualitative patterns of plant turnover are similar to those found in the manufacturing sectors of industrial countries: continual waves of small-scale entrants, many of which exit the market within the first few years of their existence. Theory suggests that heterogeneity in profit or efficiency levels and uncertainty on the part of entrants about their future ranking relative to industry norms lie behind these phenomena. In addition, turnover rates differ across industries, with high-entry industries generally characterized by high exit. At a minimum, the turnover patterns we find in the semi-industrial countries—whether measured in terms of jobs or number of plants—imply an environment with substantial resource mobility, much of it occurring among producers within the same industry

The Relationship between Productivity and Turnover

Turnover-based productivity gains can come from two basic sources. One is the continual exit of relatively inefficient producers and the simultaneous entry of producers who do better. The other source is a reallocation of market share

om inefficient toward efficient plants. These gains can be compared with the ianges in productivity within plants, which have been the focus of most studies productivity in semi-industrial countries.

To document these processes, two approaches have been used. One simply nounts to constructing output-to-labor ratios, plant by plant (see Tybout 1992). he other begins by estimating a production function describing the relation-ip between the output of a good and the inputs required to make that good ee Liu and Tybout 1996).[3] Plant-specific productivity in each year is calcu-ted as the ratio of actual output to the output predicted by the production nction, assuming a given level of inputs. Once each plant's productivity trajec-ry is calculated, it can be used to show the growth of productivity among cumbent plants that have been operating throughout the sample period and effects of turnover (entry and exit of plants). The former is simply attribut-ble to improved efficiency; the latter is attributable to productivity gaps among cumbents, entering producers (who are in their first year of operation), and iting producers (who are in their last year of operation). For example, if in-mbents are more productive than entering and exiting plants, net entry dampens oductivity growth and net exit boosts productivity. And if new producers are ore productive than the producers they replace, ongoing turnover is a steady urce of productivity gain. Algebraic details are provided in the appendix.

Some of the growth in efficiency among incumbent producers reflects im-oved productivity. But gains are also generated when resources shift from w- to high-productivity manufacturers, a shift that is generally accompanied the creation of new jobs and the destruction of old ones. We consider this rbulence arising from the reallocation of market shares to be one source of the ins in efficiency, and we distinguish it from the increase in productivity that curs within individual plants (see appendix).

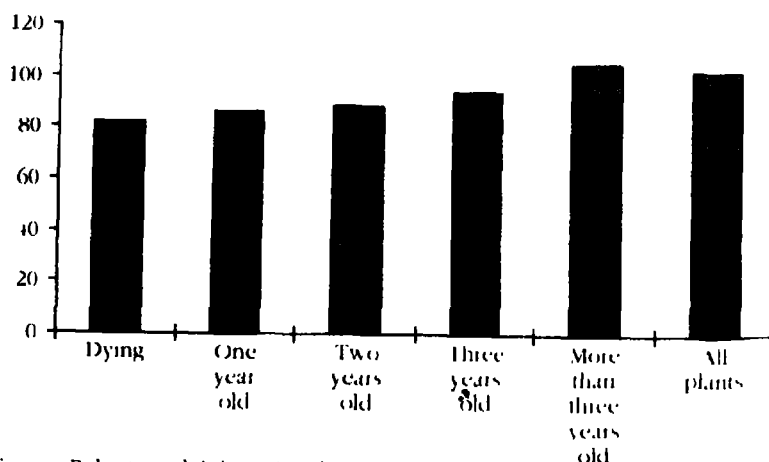
Findings on the effects of turnover in industrial countries are quite mixed. Relevant references include Baldwin and Gorecki (1991); Griliches and Regev 995; Baily, Hulten, and Campbell 1992; and Olley and Pakes 1996; for a sum-ary see Tybout (1996). For the semi-industrial countries, several basic pat-rns emerge. First, macroeconomic fluctuations can induce significant turnover d far-reaching effects on productivity. Because entry and exit rates vary dur-g the business cycle, so do the market shares of incumbent producers. During ewings incumbents lose market share because new plants enter more rapidly n incumbents fail. This pattern exerts a countercyclic influence on productiv-y because new and dying firms are typically less productive than continuing producers. Plant exits during the recession in Chile improved labor productivity ore than 1 percent, and the rapid entry of inexperienced firms during Morocco's boom period reduced labor productivity almost 2 percent (Tybout 1992).

Similarly, the productivity effect of the reallocation of market shares can be substantial in the short run, because firms do not all expand and contract pro-portionately over the course of the business cycle. In Colombia, inefficient plants shrank relatively more as the economy went into recession and recovered rela-

tively rapidly when aggregate demand rebounded. This countercyclic produ
ity effect amounted to several percentage points of efficiency gain or los
some years and industries. The *net* market share effect over the course of a
business cycle is not typically large, however.

Second, replacing dying plants with new plants also has a rather small a
age impact on productivity. The average gains during downturns are rou
offset by the losses from turnover during upswings, so most of the meas
productivity growth comes from gains in efficiency by incumbent plants. In p
this is because entering and exiting plants account for only 3 to 5 percen
production in a typical year. It is also because the productivity gap betw
plants in their first year of operation and those in their last year of operatio
small. Entering plants are only about 85 percent as productive as the industry
average, and exiting plants are roughly 80 percent as productive (figure 3)

Third, by focusing on the efficiency gap between exiting and entering plants,
substantially understates the productivity effects of turnover. As noted earlier,
average productivity of each new cohort of plants rises as it matures, reaching
dustry norms after about four years of experience (see figure 3). Thus an ente
cohort of plants eventually becomes *substantially* more productive than the col
of exiting plants it replaced, and this latter group might well have gotten worse
had not exited (Liu 1993; Griliches and Regev 1995). Although the market shar
entering and exiting plants in the transition year is small, 20 to 30 percent of
population of plants typically turns over within four years.

Figure 3. *Cohort-Specific Productivity of Manufacturing Plants in Colombia*
1982–86

Source: Roberts and Tybout (1996).

In assessing the gains from plant turnover, it is useful to ask what would have happened without any entering or departing plants? This question can only be answered with a forward-looking model of entry and exit decisions, as well as a counterfactual representation of the productivity trajectory for plants that are prevented from exiting. Developing this framework is an important topic for further research. At present we can only point to anecdotal evidence from countries that distort turnover patterns (by subsidizing or limiting entry, or by propping up inefficient producers that would otherwise exit), which suggests that the costs of such distortions are large (Pursell 1990).

Other Aspects of Productivity

In addition to the sources of productivity gains mentioned above, several others have been investigated in the research program summarized herein, including technology transfers and learning spillovers, ownership structure, scale economies, and international trade.

Technology Transfer

Harrison (1996) finds that plants owned by multinationals are typically closer to the efficient production frontier than domestically owned firms. Contrary to earlier studies based on cross sectional data, however, foreign direct investment does *not* appear to generate positive spillover effects for domestic firms in the same industry or region. At least in the short run, it appears that multinationals siphon off demand and high-quality labor from domestic competitors.

Ownership Structure

Foroutan (1996) finds that the distinction between private and public ownership is also relevant to productivity levels. In Turkey publicly owned plants are significantly less productive than privately owned plants, and they exhibit qualitatively different responses to trade liberalization. This finding is consistent with the assumption that public sector managers, lacking the disciplining influence of shareholders, pursue objectives such as job security and compensation.

Scale Economies

If there are economies of scale in production, large plants will be more efficient than small ones, so policies that influence the size of manufacturing plants also affect productivity. For example, trade development strategies may increase the size of export-oriented producers by expanding their potential market. Conversely, to the extent that economies of scale exist, the

same policies may reduce scale efficiencies in those firms that compete with imports, since these producers typically contract when trade liberalization increases import penetration in the domestic market (Rodrik 1988). As a source of productivity growth, however, such changes have probably been overemphasized relative to the other dimensions of performance. The largest plants in most industries typically have attained minimum efficient scale and these are the plants that dominate industrywide performance (Tybout and Westbrook 1996). One implication is that the computable general equilibrium models that have been used to estimate the gains from trade liberalization do not recognize differences in the size of plants within an industry and thus have often overstated the potential gains from scale economies that accompany trade liberalization (Tybout 1993).

International Trade

The degree of exposure of the domestic industry to international markets may affect productivity through other channels. Differences in productivity within an industry are typically greater in industries protected from international competition, suggesting that protection nurtures inefficiency. Higher productivity growth generally is associated with the production of tradable goods. These patterns may reflect limited access to foreign technology and expertise as well as problems acquiring imported intermediate and capital goods under protectionist trade regimes. But there are plausible alternative explanations for the negative association between protection and productivity. For example, economic models suggest that sectors with large start-up costs have relatively little turnover and tend not to sort out firms with low productivity. These sectors may also be relatively protected because they are not sectors in which the semi-industrial countries have a comparative advantage.

Conclusion

The turnover patterns we document are difficult to reconcile with the view that entry and exit primarily reflect aggregate demand fluctuations or long-term changes in technology. Instead, they seem most consistent with recent theories that emphasize the heterogeneity of producers, the uncertainty each producer faces about its ability to survive, and the constraints on turnover introduced by the sunk costs of entry and exit.

One implication of the evidence cited here is that artificial impediments that prevent failing businesses from going out of business can be very counterproductive, particularly if they are maintained over long periods of time. Mandatory severance payments or prohibitions on plant closings not only inhibit intersectoral reallocations, but also tend to discourage transfers that could lead to a more productive use of resources within an industry. It is not even clear that these

restrictive policies are useful in preserving employment. If they prevent the transfer of production to more efficient producers, they may eventually result in a smaller industrial sector. Similarly, restrictions on access to credit, equity, or other financial markets can reduce the entrance of potentially productive plants or the expansion of incumbents. Policies that reduce entry keep relatively inefficient producers in operation and slow the productive transfer of resources.

Another implication is that industrial concentration may be a very poor measure of market power; it is more likely to reflect the magnitude of sunk costs that constrain entry. Industries with high entry costs tend to have high operating profits and typically remain concentrated for long periods of time, even if firms are behaving competitively. Hence, antitrust policies designed to limit producer concentration may simply reduce efficiency.

Finally, that the cyclical component of job flows is small relative to the average level of reallocation in any year suggests that policymakers who focus on the macroeconomic causes of employment fluctuations may miss the largest source of worker transitions. Attention to the search process and associated market institutions may be a much more effective means of reducing the duration and frequency of unemployment spells and the associated efficiency losses they entail.

Appendix. Productivity Decompositions

Define $E_{it} = Y_{it}/f(v_{it}, t)$ as the efficiency of the i^{th} producers in year t, where Y_{it} is realized output, v_{it} is its input vector, and $f(v_{it}, t)$ is an estimated production function evaluated at the firm's period t input vector and the technology prevailing in period t. In some instances, capital stocks or intermediate inputs are unobservable, so $f(v_{it}, t)$ is replaced with a simple measure of factor use such as total employment. Industrywide productivity can then be written as a weighted average of the n plant-specific trajectories, $E_t = \sum_{i=1}^{n} E_{it}\theta_{it}$, where $\theta_{it} = f(v_{it}, t)/\sum_{j=1}^{n} f(v_{jt}, t)$ [illegible] is the period t market share of the i^{th} producer in terms of factor use.

To isolate turnover-based productivity growth, one can express growth in the industrywide productivity measure as the sum of three components:

$$\frac{\Delta E_t}{E_{t-1}} = \bar{\theta}_c\left(\frac{\Delta E_{ct}}{E_{t-1}}\right) + \Delta\theta_c\left[\frac{E_{bt}}{E_{t-1}} - \left(\frac{E_{ct} + E_{c,t-1}}{2E_{t-1}}\right)\right] + \left(\frac{E_{bt} - E_{d,t-1}}{E_{t-1}}\right)\left[1 - \bar{\theta}_c\right].$$

Here overbars denote averages over the periods $t-1$ and t, E_{ct} is weighted average efficiency among continuing plants (denoted by $i \in c$), E_{bt} is the weighted average efficiency among plants that enter in year t (denoted by $i \in b$), $E_{d,t-1}$ is the weighted average efficiency among plants that exit (die) after year $t-1$ (denoted by $i \in d$), and θ_c is the market share of continuing plants.

Both the second and the third term in equation A1 pick up turnover-based productivity effects. But efficiency gains attributable to resource reallocations

among incumbent plants are also potentially important and worth isolating. T this end the first right-hand side element of equation A1 can be further deco posed as:

$$
\text{(A2)} \qquad \bar{\theta}_c \frac{\Delta E_c}{E_{t-1}} = \bar{\theta}_c \left[\sum_{i \in c} \left(\frac{\bar{\theta}_i}{\theta_c} \right) \Delta E_{it} + \sum_{i \in c} \Delta \left(\frac{\theta_{it}}{\theta_c} \right) \bar{E}_i \right]
$$

where summations are only over continuing plants ($i \in c$). The first term on t right can be thought of as measuring the intraplant productivity growth effe that are the focus of representative-plant analysis. The second term—mark share reallocation effects—picks up productivity gains or losses due to size a justments among incumbents.

Notes

Mark J. Roberts is professor of economics at Pennsylvania State University and a search associate of the National Bureau of Economic Research. James R. Tybout is profe sor of economics at Georgetown University and a consultant to the World Bank. Much the material presented here is based on a World Bank research project that has been inc porated into a book entitled *Industrial Evolution in Developing Countries: Micro Patte of Turnover, Productivity, and Market Structure*, which the authors edited.

1. Kuznets (1979) observed that the common three-sector dichotomy—agriculture, manu facturing, and services—neglects all within-sector reallocations, which Kuznets sa d n have been an important reason why the productivity growth in Taiwan was poorly e plained by the structural change methodology. Syrquin (1984, p. 95) summarizes this wea ness of the traditional aggregate approach: "The estimated contributions of structural cha to growth probably underestimate the impact of resource shifts. The broad definition sectors, even in fairly disaggregated studies, hides all [intra-sectoral] factor reallo tions. . . . This is important for industrialized countries and for rapidly growing economies

2. The unit of observation in the Chilean and Colombian data sets is the manufactu plant. In the Moroccan data set, it is the firm. This is a minor point because firms ten be small, single-plant operations. In this paper we refer to the observations for t countries as manufacturing plants.

3. $Y_{it}^* = f(v_{it}, t)$, where Y^* represents the amount of output attained by the average p at the input vector v_{it} in period t. Given $f(\cdot)$, the efficiency of the i^{th} plant in year t is t imputed as $E_{it} = Y_{it} / f(v_{it}, t_{it})$, where Y_{it} is the realized output of the i^{th} plant, v_{it} is its in vector, and the denominator is a benchmark productivity level in period t.

References

Baily, Martin N., Charles Hulten, and David Campbell. 1992. "Productivity Dynamics Manufacturing Plants." *Brookings Papers on Economic Activity: Microeconomics* 1 249.

Baldwin, John, and Paul Gorecki. 1991. "Firm Entry and Exit in the Canadian Manufa turing Sector, 1970–1982." *Canadian Journal of Economics* 24:300–23.

Caballero, Ricardo J., and Mohamad L. Hammour. 1994. "The Cleansing Effect of Rece sions." *American Economic Review* 84:1350–68.

———. 1993. "Returns to Scale as a Source of Comparative Advantage." *American Economic Review: Papers and Proceedings* 83:440–44.

———. 1996. "Heterogeneity and Productivity Growth: Assessing the Evidence." In Mark J. Roberts and James R. Tybout (1996).

Tybout, James, and M. D. Westbrook. 1996. "Scale Economies as a Source of Efficiency Gains." In Mark J. Roberts and James R. Tybout (1996).

HOUSING AND INCOME DISTRIBUTION IN RUSSIA: ZHIVAGO'S LEGACY

Robert M. Buckley
Eugene N. Gurenko

> *"I'm very glad you've given up those rooms. . . We should give up still more."*
>
> Dr. Zhivago, p. 170

Th[illegible] inadequacy of housing and its effect on the quality of life go a long way [illegible] explaining many aspects of Russian life, but little data has been avail[illegible] to determine how housing affected the relative well-being of the citi[illegible]. This paper presents comprehensive data examining for the first time [illegible] effect that seventy years of a socialist housing allocation scheme had on [illegible] distribution of income. It seems clear that housing provided by the gov[illegible]ment or by employers has a value that can be measured to yield useful [illegible]ences about the distribution of income and wealth.

This article shows that housing allocation had a progressive effect on the [illegible] of income in Russia. In addition, when the imputed value of [illegible] is added to household income, the increase in income inequality [illegible] occurred in recent years is significantly reduced. The analysis concludes [illegible] discussion of how housing policy could be used to address poverty [illegible], an important aspect of the transition process.

On his return home from World War I, Boris Pasternak's fictional character Dr. Zhivago finds that he has to share his family's Moscow mansion with workers and their families. Since at least that time Russians [illegible] been sensitive to the effects of housing on living situations. The housing [illegible] and the difficulties of living in cramped quarters have been a funda[illegible] aspect of Russian life and writings about it.[1] Unfortunately, aside from [illegible] on the number of square meters of housing space produced and surveys [illegible] on interviews with emigres, little statistical material has been available to determine how housing affected the well-being and social position of Russian

World Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 19-32

citizens. Unlike Hungary, where analysts have produced a considerable body of statistical work describing housing conditions since the late 1960s (Szelenyi 1983; Hegedus 1987; and Daniel 1985), Russian researchers were not able to sift through the data to determine the role of housing in the economy.[2] The only data available from the former Soviet Union were based on the Berkeley-Duke survey of Soviet émigré households, conducted in the late 1970s. As Alexeev (1988, 1990), who examined these data, observes, this survey is not fully representative of the Soviet population. Atkinson and Micklewright (1992) suggest that data on income distribution in the former Soviet Union appeared next to alcoholism and drug addiction on the censor's list of prohibited subjects. Not until 1989 did Russia begin to count as income the overall imputed value of housing in the national income accounts (Ivanov, Rjbushkin, and Homenko 1993). And until now comprehensive data on how housing affected income distribution were not available.

The article relies on data from the Russian Longitudinal Monitoring Household Survey conducted in 1992, the year before the passage of significant reforms in housing policy. The survey, funded by the World Bank (World Bank Goskomstat 1993), was the first nationally representative household survey conducted in Russia and covered 6,128 households.[3] Our analysis of these data indicates that housing allocation had a progressive effect on the distribution of income. In addition, when the imputed value of housing is included as income the increase in income inequality recorded in recent years is significantly reduced. When housing is excluded from income, the Gini coefficient (a conventional indicator of social inequality that equals zero in the case of perfect equality and one in the case of total inequality) for Russia more than doubled between 1987–88 and 1993–94, according to a study by Milanovic (1996). That was the largest increase in the eighteen countries in the study. When housing is included in income, the increase in inequality, while still large, is significantly reduced.

Our analysis suggests that when income is adjusted for housing, the distribution of income under the Soviet regime was much more compressed than commonly thought (Bergson 1984; Gregory and Stuart 1989). Indeed, efforts to fulfill Khrushchev's dream of realizing the communist ideal by 1980, to provide workers with compensation according to need rather than performance, resulted in an extensive housing-based income transfer mechanism—a mechanism that was a fundamental part of socialist wage policy and one that significantly reduced income inequality.

To better understand housing's role in Russian welfare, we describe housing conditions and tenure patterns there and show how the inclusion of housing affects various measures of income distribution. We then consider how housing allocation affected the distribution of income and discuss the methodology used to measure imputed income. Finally, we discuss some of the policy implications of the socialist system's allocating such a large share of income through this in-kind payment, and we then speculate about how this aspect of income distribution might be affecting the transition process.

The Russian Housing Stock

The housing conditions of most of Zhivago's contemporaries in prerevolutionary Russia were dismal. A 1912 census of St. Petersburg and Moscow shows eight occupants per apartment; the comparable figure for Paris was 2.7 (Alexeev, Baker, and Westfall 1991). The average living space per urban dweller was estimated at seven square meters. For workers housing was often a bunk in a rough dormitory attached to a factory.

These substandard conditions are no longer the case for the majority of the Russian population. A massive, post-Stalin construction project—one of the largest government-sponsored investment programs ever undertaken—built nearly [illegible] million apartments in eight years (Morton 1980). Between the end of World War II and the late 1970s, the average living space per capita more than doubled, reaching 16.1 square meters in 1992 (World Bank–Goskomstat 1993).

Despite this enormous investment, the housing shortage was by no means solved. In 1992 almost 10 percent of households lived either in overcrowded communal flats or dormitories with less than eleven square meters of floor space for each household member (table 1). More than 20 percent of the residents of St. Petersburg and 17 percent of those in Moscow lived communally. For comparison, in 1974 almost 30 percent of Moscow's residents lived communally, and in 1960, 40 percent of Soviet citizens did so (Morton 1980).

Although the vigorous construction programs substantially reduced the shortage of self-contained housing units, the housing shortage—caused at least in part by the massive destruction of World War II—has never been overcome. In contrast to Western countries, where the ratio of households to the number of housing units—a simple measure of overcrowding—was about 0.9 by the early [illegible]s, the ratio in Russia stood at 1.1[illegible] and has increased since then (Alexeev, Baker, and Westfall 1991). Indeed, by 1992, according to this gross quantitative measure, housing conditions were worse in Russia than in either Poland or Al[illegible], countries noted for housing shortages (Guarda 1993).

A recent comparative study of housing conditions shows that qualitative in[illegible] of housing in Russia are far below the average of those for reforming

Table 1 *Housing Provision in Russia, 1992*

[illegible]	Russia	[illegible]	Moscow	St. Petersburg
[illegible] or sublets	2.1	1.[illegible]	0.3	0.6
Dormitories	3.0	3.9	0.9	0.0
Communal apartments	6.6	7.3	17.0	20.3
Self-contained apartments	58.9	69.1	81.8	78.8
[illegible] family houses	23.0	14.0	0.0	0.3
Private houses	5.9	3.[illegible]	0.0	0.0
Not known	0.5	0.3	0.0	0.0

Source: World Bank–Goskomstat (1993).

socialist economies in Europe (Hegedus, Tosics, and Mayo 1996). In 1992 only 44 percent of Russian households had access to running hot water; 66 percent, central heating; and 67 percent, indoor sewerage (World Bank–Goskomstat 1993). Even those who managed to escape from communal flats found their new dwellings poorly planned and constructed. As Ruble (1993, p. 234) noted: "Years of labor by residents are frequently required to correct a familiar litany of irksome deficiencies in the superficially modern Soviet high-rise apartment building of the 1990s; persistent elevator break-downs, plummeting water pressure, electrical surges, and upper-story windows that shatter in 'high winds' even though all is calm at ground level; the list goes on."

These deficiencies were embedded in the tenure structure of Russia's housing sector. The state owned and maintained almost two-thirds of the housing stock, twice the publicly owned share in Poland, almost triple Hungary's 25 percent share, and considerably larger than the 17.1 percent share in France and the [illegible] percent in the United States (Buckley, Daniel, and Thalwitz 1996). In large cities such as Moscow and St. Petersburg, the state owned almost 90 percent of the housing, in part because private housing construction was prohibited in cities with more than 100,000 residents.

The state allocated to households this publicly owned housing stock without regard to price. Indeed, for the most part, rents were set by a 1926 law, which remained nearly unchanged until 1992. Because the law held nominal rents and utility costs constant, the high inflation that occurred during the transition meant that by the end of 1992, gross housing services were essentially free.

Our approach to measuring housing income allows us to use the available data to show how this housing subsidy was distributed according to income levels and to examine how this transfer affected the distribution of income. As shown in table 2, income had little or no observable effect on distribution of housing space in 1992. The relationship between income and housing allocation was apparently random: 27.9 percent of households in the lowest income categories and 23.4 percent of households in the highest income category lived in overcrowded housing conditions—less than fourteen square meters per capita. A detailed comparison of the housing available to other income groups also fails to reveal any noticeable patterns.

Table 2. *Distribution of Urban Housing by Household Income, 1992*

Housing space per household member (square meters)	*Average monthly household income per capita (rubles)*					
	Less than 1,500	*1,501 to 3,000*	*3,001 to 4,500*	*4,501 to 6,000*	*6,001 to 7,500*	[illegible]
Fewer than 7.0	3.8	1.6	2.7	1.9	1.6	[illegible]
7.1 to 14	24.1	19.8	31.1	28.5	27.1	[illegible]
14.1 to 25	35.0	37.6	43.4	48.9	49.2	[illegible]
More than 25.1	37.1	41.0	23.1	20.7	22.1	[illegible]
Total (percent)	100	100	100	100	100	100

Source. World Bank–Goskomstat (1993).

percent went to the poorest 10 percent. Cumulatively, these results indicate that by 1992 Russia had, as Atkinson and Micklewright (1992) suggest, come to look like a market economy. The Gini coefficient for 1992 was almost identical to the United States figure for 1991 (U.S. Department of Commerce 1995).

But these figures do not include the value of subsidized housing, which was a significant component of the distribution of income. Rents were approximately zero for all but the poorest households and were only slightly higher for them. Citizens who lived in state or enterprise-owned housing paid low or no rent, in effect receiving substantial in-kind (noncash) transfers. To calculate how the benefits of unpriced housing affected the distribution of income, we derived a measure of the value of the rents and added it to each household's income to obtain a more accurate measure of total household income. We began with Hicks' (1946, p. 172) formula: "A person's income is the maximum value he can consume during a week and still be as well off at the end of the week as he was at the beginning." It is a short definition, one based on theory, and it neatly links current income from wages (or from a profit and loss statement) to the stock of wealth and the consumption associated with changes in it, such as a measure of imputed housing income.

Analyses of household income must be based on official wage statistics. If we augment that information by including the amount that would otherwise be spent on housing, we can link the notion of income to household wealth. Housing is the largest component of noncash income received by households, and it is one that the United Nations recommended be included in household income statistics.

The approach used here to impute the value of housing follows the United Nations (1968, p. 6.22) definition: "The total of owner occupied dwellings which is to be included in gross output should, in principle, be valued at rent on the market of the same facilities. It may be necessary to approximate the market rent by an estimate which should cover items such as operating, maintenance and repair outlays. . . . depreciation and mortgage interest in addition to interest on the owner's investment in the dwelling. . . ." Thus, we must first measure how much rent a household would have to pay to enjoy the same housing services provided by the state or enterprise. One way to make this notion concrete is to consider an individual who sells financial assets to buy a house with cash. Apart from such considerations as risk, liquidity, and administrative costs, if the person were as "well off" after buying the house as he was before, the imputed income (rent that the owner of the house no longer pays) would be equal to the annual interest that would have accrued on the financial assets. In terms of either the United Nations or the Hicks definition, this imputed value must be included as income if income is to be measured correctly.

We used two approaches—the first based on market value and the second on opportunity cost—to estimate the income equivalent of subsidized rent. To account for the possibility that shortages produced by the rationing system might induce households to pay more, we also used a third conceptual approach based

on the "minimum" cash compensation required to obtain adequate housing. If this approach were used, the value of imputed rents could well be higher than that estimated by the methodologies described above.

The Market Value Approach

The most commonly used approach to measuring imputed housing income is to assume it is equal to the market rental value of the housing. This estimate takes the value of imputed income to be equal to the market value of an analogous good. Although a rental housing market did not exist in Russia in 1992, it did exist in industrial nations and several developing countries. Horioka's (1994) careful analysis of household consumption and expenditure patterns in members of the Organization for Economic Cooperation and Development in 1989 found that the average household's expenditure on housing was 18.5 percent of household income, and the median was 19.2 percent. Those averages varied little from country to country.

The World Bank Housing Indicators survey (Mayo and Angel 1993) for 1991 yielded similar results. Covering housing characteristics in fifty-two cities in industrial and developing countries, this survey provides a comprehensive set of data on the housing sector, including the cost and availability of key inputs such as land, infrastructure, building materials; the regulatory environment; demographic variables; finance and subsidies; and the qualitative and quantitative features of the housing stock. These data were used to infer the median value of housing services received by a median income household in countries whose per capita gross domestic product (GDP) was similar to Russia's. This figure was computed in the form of a rent to income ratio and was drawn from a regression of rent-to-income ratios on per capita GDP in terms of purchasing power parity.

The Opportunity Cost Approach

The opportunity cost approach is the one used by Smeeding and others (1993) and Yates (1994) in measuring the imputed *net* income from owner-occupied housing in several Western countries. As noted above, the estimates of rental value for Russia take into account the provision not only of housing but also maintenance, depreciation, and utilities at essentially no cost. Hence, except for households in cooperatives, residents effectively received a 100 percent subsidy of gross rent. Based on Laidler's (1969) or Poterba's (1992) analysis, these gross rents can be converted into a rate of return that is approximately three times the 2 percent net rate of return used by Smeeding and others (1993). The gross rate of return on housing is much higher than the net real rate of return on other assets because operation and maintenance costs (about 2.2 percent) and depreciation (1 to 1.5 percent) reduce housing returns. Thus to realize a 2 percent net real rate of return requires a 5.7 percent gross rate of return (Alexeev 1990).

Renters will now have to cover the costs of operation and maintenance, multiplying their rental charges by a factor of three. Using the assumption of Smeeding and others (1993) that an average household spends about 7 percent of its income on housing, the rental of a dwelling that had cost a household 7 percent of its income will now cost it approximately 20 percent of its income.

Both approaches, then, suggest that 20 percent of income approximates the amount that the median household would have had to spend to rent the median amount of housing space it currently occupies. Because this figure may understate the amount people would actually be willing to pay, we also assumed that gross imputed housing income accounted for 25 percent of income, and considered a lower rent-to-income ratio of 15 percent (as estimated by Alexeev 1990). The direction of change in the Gini coefficients after including housing income is very similar under all three assumptions. As would be expected, the Gini coefficient is lower—that is, income inequality is more equally distributed—when housing accounts for a larger share of income, and conversely.

Because there are no recorded sales prices for housing in 1992, the technique we used ignored variations in the quality of housing. Kaganova and Malé (1994) and Pusanov (1993) recently examined these variations for St. Petersburg and Moscow, respectively. Their work shows that the assumption of invariant quality overstates the equalizing effects of housing income on the distribution of income. With variation in quality considered, the Gini coefficient for the distribution of housing in St. Petersburg rises from 23.9 to 30.7. That is, higher-income households had significantly better quality units. To compute the value of imputed gross rent for other households at various income levels we assumed that the amount of imputed income received was directly proportional to the square meters of housing provided. In other words, a family with twenty square meters of space per capita received 11 percent more in imputed housing income than residents with the eighteen square meters per capita—the median amount of housing space. Because the market rent for the median level of housing consumption per capita is equal to 20 percent of median family income, the imputed rent for a twenty-square-meter space amounts to about 22 percent of median family income. We then added this measure of income to the total income of each household in our data set.

Admittedly, this measurement values housing services solely on the basis of space. Previous estimates by the U.S. Central Intelligence Agency (1982) and Bergson (1961) rely on the same approach, however, although the latter used what he terms a "quite arbitrary" means of discounting the quality of private housing relative to public housing. In contrast Prell (1989) attempts to estimate how qualitative improvements in the housing stock in Russia, of the sort that occurred in the United States in the 1960s, may have affected the growth rate of this capital stock.

We did not attempt to make such quality adjustments for two reasons. First, we focus on one point in time rather than on how the level of investment may have changed over time as a result of unmeasured changes in quality. Second,

without more data on housing prices, we are unable to sort out how housing was distributed across households. It is by no means obvious that simple measures, such as the age or size of a building, had an effect on housing quality. Di Mao (1974), for instance, discusses the decline in quality in the larger buildings—contradicting the observations suggested by Prell (1989).

Reconsidering the Distribution of Income

Table 4 shows the cushioning effect that the imputed amount of housing income had on the distribution of income in 1992. In contrast to Yates' (1994) findings for Australia, in which inclusion of imputed housing income changed the Gini coefficient by one percentage point (from 37 to 38), the change for Russia was almost 6 percentage points. One reason for this was that Russian households did not pay for maintenance, depreciation, and utilities. The Gini for combined housing and nonhousing income is about 80 percent of that for nonhousing-related income. In other words, when these in-kind benefits are included in measures of income, there is a substantial reduction in inequality. Finally, our measure of imputed housing income accounts for about 60 percent of total income for the poorest 20 percent of the population, suggesting that these households were relatively "house rich." These households almost certainly would have preferred to receive their income in a less specific form.

Housing allocation also had significant effects on horizontal equity, that is, among persons with similar income but different in other respects (Bergson 1984). For example, we found that gender inequalities in the distribution of individual earnings were significantly offset by imputed housing subsidies because female-headed households have more housing space per capita. The inclusion of imputed housing subsidies raises the median female-to-male ratio by 7 percentage points—from 0.54 to 0.61. This effect was even more pronounced for the bottom 10 percent of the population, for which the female-to-male ratio went up by 11 percentage points (see Buckley and Gurenko 1995).

Such changes in the distributional picture are likely to have affected elements of the transition process. One possibility is that the relatively high share of housing income as a proportion of the total income of lower-income families shielded those households in their adjustment to the postcommunist depression. Milanovic

Table 4. *Inequalities in the Distribution of Combined Household Income and Income per Capita, 1985–92*

Income category	*Gini coefficient*	*Coefficient of variation*
Income per capita (including housing)	35.4	83
Income per capita (excluding housing)	41.7	99

Source: World Bank-Goskomstat (1993).

(1996), who discusses the differences between labor market adjustments duri
the Great Depression in the United States and other market economies and th
of the postcommunist depression, finds that in the former case the adjustme
took place through job cuts, while in the latter it was through wage adjustmen
In market economies, wages were broadly stable in real terms while unemplo
ment multiplied. In Russia the opposite occurred: real wages declined betwe
40 and 60 percent, and unemployment remained relatively low.

How much of this difference in labor market adjustment is attributable to t
fact that much of the income of lower-wage workers was unaffected by wa
cuts? Did enterprises in Russia, which had traditionally provided most of t
social safety net, respond in the way they did because employees' househo
income would not be affected by such cuts? Conversely, does the illiquidity
the imputed housing income affect the adjustment of labor markets? One rece
study of Poland, for example, estimates that as much as 25 percent of its unem
ployment rate in 1992 occurred because workers could not find housing ne
the available jobs (Coricelli, Hagemejer, and Rybinski 1995). Although we
not know the answers to these questions, the adjustment is a part of the ec
nomic landscape that should not be ignored.

As shown in table 4, the distribution of combined income (including housin
was substantially more compressed than the distribution of income derived fro
wage data or total earnings in the techniques used by Atkinson and Micklewrig
(1992). Moreover, one has the sense that if such information could have be
spliced onto the 1989 comparisons, the Soviet Union would not have looked
nearly like the Western economies but instead more like the socialist regime th
it was.

Conclusion

The distribution of housing in Soviet Russia reduced income inequali
and provided a strong cushion against the consequences of the transiti
The effects were similar to those described by Smeeding and others (199
for the OECD economies, but that study found that housing subsidies repr
sented the smallest share of in-kind compensation. Moreover, in their find
ings the total share of all nonwage compensation, 21 percent, was only slightl
more than our measure of imputed income from housing. In Russia, in con
trast, it can be inferred from Alm and Sjoquist (1995) that housing accounted
for the largest share of in-kind compensation. They found that expenditure
on housing-related maintenance alone, which are just a small share of in
kind housing compensation, accounted for nearly half the financial obliga
tions for social services that firms were now transferring to local govern
ments. Housing is certainly not an income source that can be omitted from
consideration (as does Kakwani 1995) on the grounds that the rents charged
by the state are low.

An important manifestation of the cushioning provided by housing can also be seen by considering housing's role in national wealth, particularly in light of the greater dispersion of the distribution of wealth relative to that of income. In France and the United States, for example, housing accounts for one-half and one-third, respectively, of household wealth. In both countries, wealth is far more dispersed than income (Kessler and Wolff 1991). In a society like Russia's, in which inflation has eliminated most savings of the household sector, imputed income from tangible wealth such as housing is likely to be an even larger component of wealth than it is in market economies. And because of high rates of inflation, nonresidential wealth is likely to have been more widely redistributed than in market economies. As a result, the almost "give away" privatization of housing that began in Russia in 1993 gave households at least some savings to help cushion the costs of the transition. This program was an important way of permitting households to exploit the distributional benefits of the old system, a step that has not been pursued nearly as aggressively in many other reforming economies. Because the distribution of housing income in Russia is so much less dispersed than is wage income, while at the same time the distribution of nonresidential wealth is almost certainly much more dispersed, housing's privatization had a strongly progressive effect on the distribution of wealth.

The privatization of housing may also help address poverty concerns. For example, in 1992 the elderly in the lowest income quartile had less than one-seventh of the income of young households in the highest income quartile. But they had five square meters more housing space per person. Because up to 30 percent of the elderly were below the poverty line, being able to liquidate this additional housing wealth could go a long way toward addressing their poverty problem.

Devising means by which the poor and the elderly can gain access to this housing wealth is by no means a simple task. Complicated questions of property rights, registration, and enforcement as well as such issues as defining condominium rights and responsibilities are fundamental management issues that beg for simple resolution. Nevertheless, at the very least, this source of income should not be ignored, as it has been in recent analyses (see, for example, Barr 1993). We recognize that our data could be significantly improved, particularly as market transactions permit better calculations, but given the scale of the effects involved, the evidence supports a 1977 United Nations recommendation that housing income be included in analyses of income distribution.

The housing policies that caught Dr. Zhivago's attention more than seventy years ago continue to affect the basic fabric of life in Russia. Decisions with respect to marriage, children, profession, and job opportunities are no doubt affected by housing considerations. The impact of housing policies on so many aspects of economic behavior makes social compensation systems more difficult to reform. Thus, at a minimum, the former socialist economies should place a greater emphasis on integrating housing decisions into broader measures of economic activity.

Notes

Robert M. Buckley is with the Operations Evaluation Department of the World Ban and Eugene N. Gurenko is with Planning and Development Collaborative (PADCO), Wa ington, D.C. This article was initiated as a component of a World Bank research progra The authors would like to thank Michael Alexeev, Patricia Annez, Lev Freinkman, Jo Hegedus, Stephen Mayo, and Blair Ruble for their valuable assistance.

1. Pasternak's interest in housing has been a constant theme of Russian literature Se for example, Brodsky's "In a Room and a Half," in *Less than One: Selected Essays* (19 and the chapter called "Housing the People," in *Khrushchev Remembers: The Last Tes ment* (1980).

2. Szelenyi was exiled from Hungary when his studies, conducted in the late 1960s a early 1970s, showed that the Hungarian mechanism for allocating housing resulted regressive distribution of benefits. (His work was published in English in 1983)

3. The Gini coefficient for 1985 was 25.6; for 1989, 28 9; and for 1992, 41 7 S Atkinson and Micklewright (1992) for 1985 and 1989 figures; World Bank–Goskoms (1993) for 1992 estimates based on the same definition

4. A detailed description of the data set is available upon request

References

The word "processed" describes informally reproduced works that may not be comm available through library systems

Alexeev, Michael. 1988. "The Effect of Housing Allocation on Social Inequality A So Perspective." *Journal of Comparative Economics* 12:228 34

———. 1990. "Distribution of Housing Subsidies in the USSR, with some Soviet-Hun Comparisons." *Comparative Economic Studies* 32(3) 138 57

Alexeev Michael, Lee Baker, and Michael Westfall 1991 "Overview of the So Housing Sector." Paper prepared by the Planning and Development Collabo Washington, D C. for USAID Office of Housing and Urban Program Processed

Alm, James, and David Sjoquist 1995 "Social Services and the Fiscal Burden in Rus *Comparative Economic Systems* 37(4):19 30

Atkinson, Allan, and J. Micklewright 1992 *Economic Transformation in Eastern E and the Distribution of Income* Cambridge, U.K Cambridge University Press

Barr, Nicholas. 1993. *Income Transfers and the Social Safety Net in Russia* Stud Economies in Transformation 4. Washington, D C World Bank

Bergson, Abraham. 1961. *The Real National Income of Soviet Russia Since 19* bridge, Mass.: Harvard University Press

Brodsky, Joseph. 1986. *Less than One. Selected Essays* New York Farrar, Straus, G

Buckley, Robert M., and Eugene N Gurenko 1995 "Distribution of Income and H in the Period of Economic Transition in Russia: Discussion of Data and Further A sis." World Bank, Operations Evaluation Department, Washington, D.C Process

Buckley, Robert M., Zsuzsa Daniel, and Margret Thalwitz 1996 "The Cost and T parency of Housing Subsidies in Socialist Economies " In Robert M Buckley, ed., *Hou Finance in Developing Countries* London: Macmillan Press

Coricelli, Fabrizio, Krzysztof Hagemejer, and Krzysztof Rybinski 1995 "Country St ies: Poland." In Simon Commander and Fabrizio Coricelli, eds., *Unemployment Re structuring and the Labor Market in Eastern Europe and Russia.* EDI Development Stu ies. Washington, D.C.: World Bank.

Daniel, Zsuzsa. 1985. "The Effect of Housing Allocation on Social Inequality in Hu gary." *Journal of Comparative Economics* 9:391–409.

dies for Health, Education, and Housing." *Review of Income and Wealth* (September):229–56.

Szelenyi, Ivan. 1983. *Urban Inequalities under State Socialism.* Oxford: Oxford University Press.

United Nations. 1968. *A System of National Accounts.* Studies in Methods, Series F, No. 2, Rev. 3. Department of Economic and Social Affairs. New York.

———. 1977. *Provisional Guidelines on Statistics of the Distribution of Income, Consumption, and Accumulation of Households.* Studies in Methods, Series M, No. 61. New York.

U.S. Central Intelligence Agency. 1982. *Soviet Statistics on Capital Formation.* SOV 82-1000093, August.

U.S. Department of Commerce. 1995. Personal communication. Research Department. Washington, D.C.

World Bank–Goskomstat. 1993. *Russian Longitudinal Monitoring Household Survey.* Washington, D.C.

Yates, Judith. 1994. "Imputed Rent and Income Distribution." *Review of Income and Wealth* 40(1, March):43–66.

AN EXCHANGE ON PROJECT EVALUATION

Because of the importance of investment to development, the economic appraisal of projects has long attracted the attention of development theorists and practitioners alike. The subject has recently been enjoying a revival of interest, as several studies revealed a decline in the quality of economic analysis used, both in the World Bank and elsewhere. The following three articles reflect some of the views expressed in the recent resurgence of interest in project analysis. We hope that these articles will stimulate further contributions, as the development community grapples with this important and evolving subject.

BEYOND RATE OF RETURN: REORIENTING PROJECT APPRAISAL

Shantayanan Devarajan
Lyn Squire
Sethaput Suthiwart-Narueput

Traditional approaches to project appraisal fail in practice to address two fundamental questions: whether a project belongs in the public or the private sector; and what effect any external assistance associated with the project has on the country's development. The first issue is of general interest to all national policymakers and international donors. If the government provides a good or service that would otherwise have been provided by the private sector, the net contribution of the public project could be low. The second issue is of particular concern to donors. If financial resources are fungible, the project being appraised might well have been undertaken without external financing. In this case, donor funds are actually financing some other unappraised project. Both cases argue for a shift in the emphasis of project evaluation away from a concern with precise rate-of-return calculations and toward broader sectoral analyses and public expenditure reviews. In this context, three areas critical for proper project appraisal include a consideration of the rationale for public intervention, the fiscal impact of the project, and the fungibility of external assistance.

A quarter century ago, several groups of economists (Little and Mirrlees 1969, 1974; Dasgupta, Marglin, and Sen 1972; Harberger 1973; Squire and van der Tak 1975) developed methods for appraising investment projects financed from public revenues. Extending the principles used by the private sector in making investment decisions, these economists advocated the use of "social cost-benefit analysis," a test that weighs a public project's costs and benefits in terms of its contribution to national (social) welfare. If the project's social benefits exceeded its social costs, the recommendation was that the government should undertake the project.

Although there are important differences in the various methods for evaluating the costs and benefits, they have at least three elements in common. First,

The World Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 35–46

they emphasize that a project's inputs and outputs should not necessarily be valued at current market prices because the market price may not reflect the social opportunity cost of the resource—that is, its cost in forgone benefits to society. Say, for example, that a project hires an unemployed worker. Even though his wage is a cost to the project, it does not represent the social cost, which would be the supply price of labor (that is, the price at which a person would be willing to work). When there is unemployment this usually means a price (or wage) that is below the prevailing—or actual—wage. Thus, rather than use market wages as the cost of labor, for example, the evaluator is instructed to use a set of "shadow prices" that reflect the social opportunity cost of the project's inputs and outputs.

Second, policymakers should evaluate every component of a project relative to a counterfactual—that is, what would have happened without the project. The example above also illustrates this principle: because the worker in this case was unemployed, no national output is forgone when he is drawn into the project. Finally, the methodology dictates that because the project's benefits and costs occur at different points in time, they should be combined in some summary statistic, such as the project's net present value or internal rate of return.

Today, a charitable assessment would have to acknowledge that the practical application of these principles has been limited. Governments and international agencies (some of which developed social cost-benefit analysis) use the techniques only sporadically. In a retrospective paper Little and Mirrlees (1991, p. 376) concluded "that the extent to which [social cost-benefit analysis is] used and [has] real influence is not great, even in the World Bank."

Although there are many reasons for this decline in the use of social cost-benefit analysis, Little and Mirrlees (1991) suggest four: the growth of nonproject lending; internal incentives in lending institutions; new concerns such as poverty, women, environment, and "sustainability"; and the complexity of the methodology. In addition, the traditional approach may often fail to address the fundamental questions of concern to policymakers and donors today. Among these questions two of the most important are whether a project belongs in the public or private sector and what the effect has been on development of the external assistance (if any) associated with the project?

The first question is of general interest to both national policymakers and international donors. The world has changed substantially in the past twenty-five years. Governments and donor agencies are now reconsidering the role of the state. Instead of asking if the project generates a positive net social benefit, governments and agencies are asking if there is a rationale for public provision of that good or service. Whether the standard approaches to project analysis are suited to answering the latter question is not clear. For instance, a recurring example in Little and Mirrlees' (1974) book is an industrial project in India—a project whose rationale for public intervention would be dubious today. By contrast, projects with a strong public-sector rationale, such as vector control and immunization (described in Hammer's companion article in this volume), are

likely to be those with large positive externalities (that is, their value to the community is greater than their price). The resulting improvement in welfare, although apparent, is difficult to quantify in net present value or rate-of-return calculations.[1]

The second issue is of particular concern to donors. If financial resources are fungible, it is unclear what a project's rate of return conveys about the loan's effect on development. The project being appraised might well have been undertaken without external financing. In this case, the donor's funds are actually financing some other project that would not have been carried out otherwise.

What is the role of project analysis in a world in which the public-private boundary is the relevant issue? This article shows that the principles underlying social cost-benefit analysis can incorporate this dimension but that the technique must be modified. Because the projects under consideration by policymakers and donors are public projects, an appropriate counterfactual is what the private sector would have provided in the absence of the project. What is required is sectoral anlysis to identify market failures (that is, instances where the market over- or under-provides goods or services relative to the socially desirable levels) that warrant public intervention. Reorienting project appraisals in this direction also leads policymakers to focus on the project's fiscal impact. The article addresses the problems that arise with standard net present value and rate-of-return calculations when financial resources are fungible and argues that policymakers should not rely on such techniques. A more appropriate analysis is a review of public expenditure programs, both to improve the quality of projects and to ensure that the impact of foreign aid on development is favorable.

The Rationale for—and Cost of—Public Provision

Traditional cost-benefit analysis, as presented by Little and Mirrlees and many others, [illegible] addressed the following question: Will the project under consideration yield a net benefit to the economy? This is an important question. No one wants to invest in projects that impede overall development. But the answer to this question says nothing about whether the project ought to be in the public or private sector.

Twenty-five years ago that concern may not have been so great. At the time governments were expanding public investment rapidly, and much of that investment was in industrial sectors. Indeed, the first version of Little and Mirrlees' work, published in 1969, was titled *Manual of Industrial Project Analysis*. But now countries worldwide are redrawing the boundary between the public and private sectors and paying more attention to whether or not a project ought to be in the public sector.

The *principles* underlying the manuals on project evaluation can be applied to the changed circumstances and can help to formulate a new set of questions about the appropriate role of government. In particular, all the approaches to

cost-benefit analysis require the project evaluator to specify the *counterfactual*. This principle is perfectly general; it applies even when the alternative to public provision is a private project.

To illustrate, assume that the government is contemplating an investment that produces private goods—a shoe factory, say. One possible judgment regarding the counterfactual is that, in the absence of the public-sector project, nothing would have happened. In this case, the analyst should focus on a comparison of the costs incurred by the project and the benefits it is expected to yield. Assuming that the evaluation points to a positive net present value (that is, the benefits are higher than the costs), the decision would be made to go ahead with the project. But the appropriate counterfactual might be that the private sector would have produced the shoes anyway (assuming the enterprise is profitable). In this case, the relevant magnitude is the *net* contribution of the government shoe factory. The net present value of the public-sector project over and above that of the private-sector project (evaluated at shadow prices) may well be zero.

Specifying the counterfactual is not always easy. The relevant counterfactual to public provision includes the private market outcome considering taxes, subsidies, and regulations. But some principles can be invoked. First, if the project is producing a private good that is profitable at market prices, there is good reason to believe that the private sector will undertake it. In this case there is no advantage to public provision, and the point can probably be established without a serious evaluation. Second, at the other extreme is the case of pure public goods. Here there is no prospect of private provision and hence no need to worry about a private-sector counterfactual. But pure public goods—defense, for example—are relatively rare.

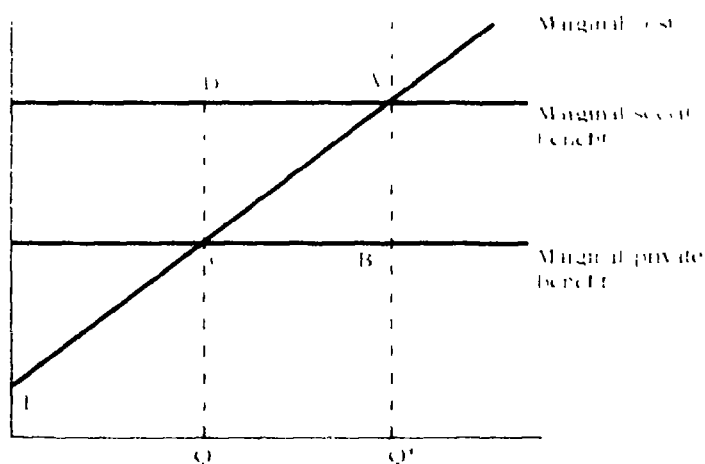
The third possibility is the need to redistribute income to the poor or correct for market failures, such as externalities and public goods. In such areas private markets will not yield a socially desirable outcome, and a case can be made for government intervention (even though the private-sector counterfactual will not necessarily be zero). Because many projects are likely to fall under this category, we consider it in more detail.

Consider the case of a product with a positive externality—a benefit from producing a good or service that is not fully captured in the product's price. The private sector will provide some of the good, but because producers are not paid the full social value of it, they will produce less than the socially optimal amount. In figure 1, the positive externality is shown as the marginal social benefit, which describes how much society is willing to pay for each unit of the good. This amount is greater than the marginal private benefit—or private demand, which is the price that individuals are willing to pay. The marginal social and private benefit lines are drawn as perfectly elastic for ease of exposition. An example would be secondary education, which many would claim has significant positive externalities, although in many countries private secondary schools exist alongside public ones. In figure 1 the private sector would provide up to Q of the good

on its own. The socially optimal amount, however, is Q^*. (In the shoe factory example referred to earlier, the marginal social benefit and marginal private benefit curves coincide, so that $Q = Q^*$.)

In evaluating projects such as these, the analyst needs to establish three points. First, public provision should result in a greater supply of the good than would have occurred with just private-sector provision (that is, overall supply should exceed Q). It is also possible that private provision up to Q is not forthcoming because, for example, of capital-market imperfections. But if that were the case, direct intervention in the capital markets would be better than addressing the problem indirectly through public provision. It is important to ascertain precisely why private supply is not forthcoming. The current level produced by the private sector could also represent that amount of output that can be profitably produced and sold given the costs imposed by government-induced distortions. In that case the appropriate comparison should be between the net benefits of the project and the costs of removing those distortions. If the state displaces private investment, it would simply be "crowding out" private providers. The amount produced by the private sector in the absence of the project (the counterfactual) would be equal to the amount produced by the project (assuming costs are the same), and thus the net present value of the public-sector project would be zero.

Second, the project should not result in total supply of the good beyond Q^* (the point at which marginal costs exceed marginal social benefits). Thus, the relevant range for public provision in this sector is between Q and Q^*.

Figure 1. *Public Provision and the Private-Sector Counterfactual*

Note: Each cost (benefit) is the cost of (benefit from) an additional unit of output.
Source: Authors' calculations.

Third, although the presence of spillovers and market failure may justify public *intervention*, it should not be presumed that public *provision* is required. The relevant counterfactual to public provision could well be the private market outcome under appropriate taxes, subsidies, or regulation. In figure 1, for example, an alternative to public provision up to the optimal quantity Q^* would be a subsidy equal to height AB, the size of the positive externality at Q^*.

Figure 1 is drawn as though the exact size of the externality were known so that the optimal amount the public sector should provide—Q^*—could be determined. In reality, the true value of the externality will rarely be known. Indeed, that is true of most justifications for public intervention, whether the rationale is a market failure or redistribution. Until better estimates of externalities are available, there is no real solution to this problem. Thus the risk of over- or under-supply cannot be eliminated.

The *costs* of public-sector intervention, however, can be measured much more easily. They can be approximated by calculating the additional burden the project imposes on the budget when consumers are charged the full amount they are willing to pay. In figure 1 the area under the marginal cost between C and Q^* is the total additional cost, but consumers are paying $BCQQ^*$ to the government; thus the additional cost of the project is shown by the area ABC. Of course, this area can be identified only because the size of the externality is known. Assume that the size of the externality is unknown—that is, the marginal social benefit line in figure 1 cannot be located. In fact, that line could be anywhere above the line depicting the marginal private benefit. If the government still decides to supply Q^*, then the area ABC can be interpreted as the additional cost incurred in order to realize the (unknown) benefit implied by the existence of the externality. Such cost calculations can provide a useful "reality check" on proposed interventions. Whatever the true size of the external benefits, the government must judge that *at a minimum* the external benefit exceeds this cost for the intervention to be worth undertaking. To illustrate, CNCA (Caisse Nationale du Credit Agricole), a development bank serving rural areas in Morocco, received subsidies that could conceivably be justified on the grounds that the bank operated in an underserved rural credit market and reached poor people. Although these benefits are hard to quantify, assessing the cost of the subsidy is one way to ask whether this subsidy is a good use of scarce public resources and to think about alternative uses. In this case CNCA's annual subsidy amounted to about 20 percent of the recurrent budget for primary education and 160 percent of the recurrent budget for basic health care. And this in a country where social indicators are quite unsatisfactory: primary enrollment is around 70 percent, and under-five mortality is about eighty deaths per thousand live births.

It should be emphasized that while Q is an unobserved counterfactual, it is routinely assessed implicitly during project appraisal. Most project appraisals include calculations of both the financial net present value (using market prices) and the economic net present value (using shadow prices). A positive financial

net present value strongly indicates that the private sector could undertake the project (because it generates profits).[3] Such a positive value is thus an argument *against* public-sector investment in the project. (This does not mean that the project should necessarily be undertaken from a social standpoint, simply that it should not be undertaken by the public sector. A private project that generated negative externalities—pollution, for example—could be privately profitable but socially undesirable.) Note that this conclusion is exactly the opposite of current practice where a high and positive financial net present value is used to *justify* a public sector project.

An estimate of what consumers are willing to pay provides an indication of how much the private sector would have provided, because a private company should be willing to provide goods as long as it is profitable to do so. An appraisal of the Leyte Luzon Geothermal Project in the Philippines, a $1.3 billion project, illustrates the idea. Consumers were willing to pay 6.8 cents a kilowatt-hour (kWh) for electricity, based on current bulk energy rates in Luzon, compared with an estimated long run marginal cost of 5.2 to 5.8 cents/kWh to operate the project. Because consumers were willing to pay more than the cost to produce the electricity, it would have been profitable for a private company to undertake the investment. Public provision here crowded out at least this quantity of private provision.

The same argument holds for projects that aim to redistribute income. For these projects the government charges less than consumers are willing to pay and thus incurs a fiscal cost. Although the incidence of such redistributive transfers may be easier to quantify than externalities are, the government still needs to gauge whether the value of the redistribution outweighs its fiscal cost.

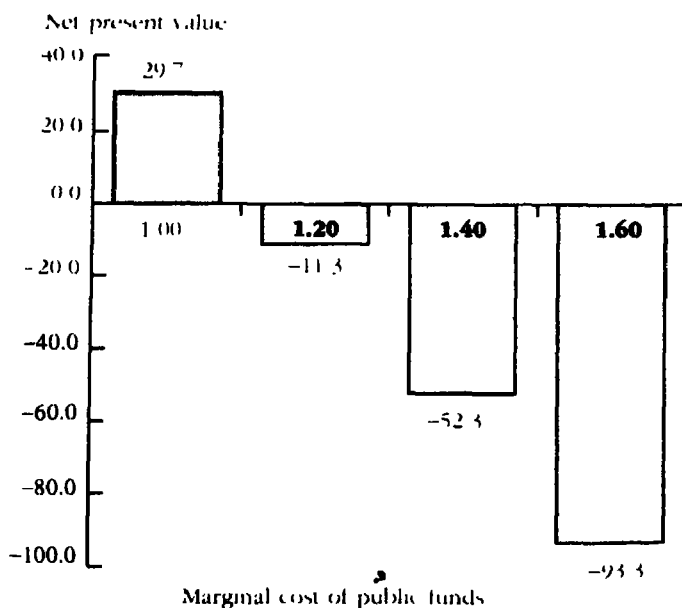
As the public sector shifts the composition of projects in favor of those with a clear public sector rationale—on either market failure or redistributive grounds—these projects will usually imply a subsidy. To put the same point differently, these projects place a burden on the budget. Another change in circumstances, more accurately, in the appreciation of what conditions support development, is the recognition that macroeconomic stability is an essential prerequisite. Since a prudent fiscal policy is central to macroeconomic stability, it follows that projects that require government funding need to be reviewed with care. Fiscal balance requires that the government recoup the costs of these projects through some other tax instrument, which in turn will introduce distortionary costs somewhere in the economy. These costs should be included in the evaluation of the project. If they are omitted, and public costs and private benefits are weighted equally, the net present value of these projects will be systematically overestimated (Squire 1989). Although policymakers do not have a precise measure of the marginal cost of funds for most countries, the net impact of the project on the government's budget—a minimum measure of the true cost—is still worth showing. Because of the uncertainty surrounding the marginal cost of public funds, it is useful to reestimate the project's net present value for a range of values of the marginal cost of funds (box 1).

Box 1. *Revisiting the Marginal Cost of Public Funds*

How does the government generate funds to cover the fiscal cost of a project? Even if the funds are borrowed, ultimately the government's only source of revenue is the domestic tax base. Every tax instrument, however, imposes a cost on the economy because it creates a distortion. This additional cost should be applied to the project.

How much does it cost? Ballard, Shoven and Whalley (1985) estimate that it costs seventeen to fifty-six cents in the United States to raise a dollar of extra revenue (yielding a marginal cost of public funds of $1.17–$1.56). This suggests that public projects should produce marginal benefits of more than $1.17 per dollar of cost. In developing countries, the marginal cost of funds is likely to be even higher to the extent that these countries have access to a more limited set of tax instruments (trade taxes, for example), which are highly distorting.

To ensure that public-sector projects recover as much of their costs as possible from the private sector, an appropriate pricing strategy is important. Figure B-1 uses the geothermal project in the Philippines to show how a project's net present value changes with different pricing policies and different levels in the marginal cost of funds. If the electricity is sold at 80 percent of the market value and there is no distortion associated with increased taxation (that is, the marginal cost of funds is one), the net present value for the project is $29.7 million. But with even a small premium of twenty cents to raise each dollar, the project's net present value becomes *negative*. With a marginal cost of funds of 1.6 (roughly the upper range estimated for the United States), the project's net present value drops to –$93.3 million.

Figure B-1. *Leyte-Luzon Geothermal Project: Project Net Present Value at Discounted Price for Electricity*

Note: Assumes 80 percent of market rate for electricity charges and 10 percent for discount rate.
Source: Project documents; author calculations.

The Fungibility of Aid

Project-specific appraisals can (at best) only assess the project's rate of return or its acceptability. This approach is problematical for two reasons, both of which are important for multilateral lending agencies and donors interested in the impact not only of aid-financed projects but also of aid itself. First, as noted earlier, financial resources are fungible at least to some extent (Feyzioglu, Swaroop, and Zhu 1995; Pack and Pack 1990, 1993). It is unlikely that the projects evaluated by the World Bank, for example, are so marginal that they would otherwise not have been carried out. For the ninety-nine projects evaluated in 1993, the World Bank (1994) found an average economic rate of return of 21 percent—a return too high to indicate marginal projects. Second, even if the project would not have been undertaken without external funding, there is no guarantee that it was the *best* of all the projects under consideration. Yet that is the relevant question.

One practicable approach is to require sectorwide reviews before project-specific appraisals and financing decisions are made. For example, Humplick and Paterson (1994), who studied the infrastructure sector in Peru, calculated the economic rate of return of expanding each of the major road links in the country. The results varied widely. The report recommended funding only those projects with a rate of return above 12 percent. On this basis several road expansion projects were dropped, cutting $275 million from the government's road construction program. Donor financed road projects may have had high appraised rates of return, but if these projects were already included in the government's prospective investment program, the development impact of donor financing would have allowed funds to be shifted to one of the projects with a rate of return below the 12 percent cutoff rate.

These reviews actually go beyond setting a good foundation for subsequent project appraisal. They also improve the overall quality of the sectoral investment program. Moreover, if the result is satisfactory, the specific project financed and appraised by the donor agency becomes less important. The donor could have more impact on development by associating itself with that project where its knowledge and technical expertise is likely to be of most value. Alternatively, the agency could finance a "time-slice" of a specified expenditure program, an option that is receiving greater attention (Harrold and Associates, 1996). (Time-slice lending finances a certain percentage of public expenditures for a certain number of years rather than the whole project.)

Sectoral expenditure reviews can also shed light on the hypothetical no-program (or counterfactual) state by identifying areas where the private sector already provides, or can provide, the goods and services in question. For instance, a review of public expenditures in Malaysia (World Bank 1992) noted that 62 percent of the Ministry of Health's funds went to medical care (mostly private goods) and only 23 percent to public health, although the latter clearly had a higher marginal impact on health.[4] Although sectoral expenditure reviews do not provide data on

the marginal cost of public funds, public expenditure reviews regularly conducted by the World Bank that cover a government's entire budget could in principle do so. At the very least, public expenditure reviews should identify those countries in which the gains from applying the marginal cost of funds in project evaluation—or the losses from not doing so—are the greatest.

Conclusion

Altogether governments in developing countries typically spend about 4 percent of gross domestic product, or a total of about $190 billion on public investments every year. Overseas development assistance runs at about $45 billion annually. Even marginal improvements in assessing investment projects could therefore have very high payoffs. At the same time, the analytical skills to undertake careful evaluations are limited and the information required is often lacking.

To some extent, the changed circumstances justify the relatively limited attention that government and donors have paid to the appraisal process (World Bank 1995). In this context, how should the scarce analytical resources that are available be allocated to improve the quality of projects and achieve the maximum impact on development?

Public Expenditure Reviews

Governments and donors should routinely review the entire public expenditure program or its sectoral components *before* embarking on the appraisal and financing of specific projects. Governments would benefit because such a review would improve the overall quality of the public investment program; donors would benefit because the program would provide some assurance that aid flows were being well utilized even when financial resources were fungible.

Once an expenditure review is in hand, aid programs should be designed to ensure that the recommendations of the review are implemented and that financial assistance is conditioned on a satisfactory program of public spending. The choice of lending instrument—single project, sectoral investment loan, or general budgetary support—should be made according to which vehicle contributes most effectively to the objective. Where the investment program is less than satisfactory, donors would be well advised to focus their efforts on technical assistance and to limit their financial assistance to projects that are likely to be nonfungible.

Project Evaluation

Whether in the context of a single project or a public expenditure review, project evaluation still has a vital role to play. But the traditional tool analysts use—the rate of return—will be less and less relevant given the nature of the

projects that governments and donors will be pursuing. In these new circumstances, two questions should be routinely addressed in evaluating every project, even when it is not possible to measure all the benefits, a characteristic that is likely to describe a growing number of public investment programs.

First, what is the rationale for the public sector's involvement? Since governments and donors are still financing projects that appear to be producing private goods, a greater effort should be made to establish the rationale for public provision. Policymakers should routinely assess private-sector alternatives, including improved regulatory or price (tax-*cum*-subsidy) policies. Even where the rationale for public ownership is clear, the objective should be to ensure that total supply is greater than would be the case with just private provision.

Second, what is the cost of public provision? If countries and donors systematically implement the preceding recommendation, the net impact on the budget will be negative. Where this fiscal cost arises from an expansion in supply beyond what would have been forthcoming from the private sector, it represents the price that society has to pay to reap the benefits underlying the rationale for public intervention. If the government is not charging the maximum amount that the private sector is willing to pay, there is an additional fiscal cost—a transfer. Both the expansion and the transfer constitute additional burdens on the budget. To the extent that governments have to rely on (distortionary) taxation, raising the required revenue will entail real costs. These costs, as well as the marginal cost of public funds, need to be incorporated in project appraisal whenever possible.

Are our proposals feasible? Technically, yes. For example, the rationale for public sector provision can be identified at the same time as the public expenditure review is conducted or early in the project cycle, and without amassing all the information usually associated with a full blown cost-benefit analysis. Similarly, analysts should be able to provide reasonable estimates of the fiscal costs of projects without knowing all the benefits. Finally, both of these tasks can be undertaken for all projects.

Notes

1. Shantayanan Devarajan is division chief, Lyn Squire is director, and Sethaput Suthiwart-Narueput is an economist in the World Bank's Policy Research Department. The authors thank Mark Baird, Pedro Belli, Jeffrey Hammer, Anandarup Ray, and several anonymous referees for their insightful comments on an earlier draft.

2. In this article rate of return refers to both rate of return and net present value calculations.

3. Note that this is true for any intervention up to Q^*. At Q^*, for example, the additional cost *ABC* is strictly less than the size of external benefits beyond the market outcome *ABCD*. For interventions beyond Q^*, this may no longer be the case.

4. The qualification that capital market imperfections may prevent the private sector from undertaking the project may apply here, as does the response, namely, that the appropriate intervention would be to remove the capital-market imperfection.

4. Analyses of infant mortality showed that expenditures on safe water and immunization had a much higher effect than expenditures on government-employed doctors (World Bank 1992).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Ballard, Charles, John Shoven, and John Whalley. 1985. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *American Economic Review* 75(1):128–38.

Dasgupta, Partha, Stephen Marglin, and Amartya K. Sen. 1972. *Guidelines for Project Evaluation*. New York: United Nations Industrial Development Organization.

Feyzioglu, Tarhan, Vinaya Swaroop, and Min Zhu. 1995. "The Impact of Foreign Aid on Public Spending." World Bank, Policy Research Department, Washington, D.C. Processed.

Harberger, Arnold C. 1973. *Project Evaluation: Collected Papers*. New York: Macmillan.

Harrold, Peter, and Associates. 1996. *The Broad Sector Approach to Investment Lending: Sector Investment Programs*. World Bank Discussion Paper 323. Washington, D.C.: World Bank.

Humplick, Frannie, and William Paterson. 1994. "Peru: Infrastructure Public Expenditure Review." World Bank, Latin America and the Caribbean Region, Washington, D.C. Processed.

Little, I. M. D., and J. A. Mirrlees. 1969. *Manual of Industrial Project Analysis*. Paris: OECD Development Centre.

———. 1974. *Project Appraisal and Planning*. London: Heinemann.

———. 1991. "Project Appraisal and Planning Twenty Years On." In *Proceedings of the World Bank Annual Conference on Development Economics 1990*. Washington, D.C.: World Bank.

Pack, Howard, and Janet Rothenberg Pack. 1990. "Is Foreign Aid Fungible? The Case of Indonesia." *Economic Journal* 100(399, March): 188–94.

———. 1993. "Foreign Aid and the Question of Fungibility." *Review of Economics and Statistics* 75(2, May): 258–65.

Squire, Lyn. 1989. "Project Evaluation in Theory and Practice." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 2. New York: Elsevier Science Publishing Company.

Squire, Lyn, and Herman G. van der Tak. 1975. *Economic Analysis of Projects*. Baltimore, Md.: Johns Hopkins University Press for the World Bank.

World Bank. 1992. "Malaysia: Fiscal Reform for Stable Growth." Report 10120-MA. Asia Region, Washington, D.C. Processed.

———. 1994. *1993 Evaluation Results*. Washington, D.C.

———. 1995. "A Review of the Quality of Economic Analysis in Staff Appraisal Reports for Projects Approved in 1993." Operations Evaluation Department, Washington, D.C. Processed.

ECONOMIC ANALYSIS FOR HEALTH PROJECTS

Jeffrey S. Hammer

This paper applies to the health sector a method of project analysis advocated recently by Devarajan, Squire, and Suthiwart-Narueput. A health project evaluation should establish a firm justification for public involvement, establish the counterfactual—what would happen with and without the project; and determine the fiscal effect of the project and the appropriate levels of fees in conjunction with project evaluation. The evaluation should also acknowledge the fungibility of project resources and examine the incentives both for high-level public servants to shift government resources away from project-funded activities to those that have not been evaluated and for lower-level contractors and civil servants to provide good or bad service. Market failures in health services and insurance markets should serve as a starting point for economic analysis, not as a reason to ignore economics in health projects. Project outputs should be predicted after taking into account the reaction of consumers and providers in the private sector as well as market structures of supply, demand, and equilibrium for health services.

A recent article summarized elsewhere in this volume, Devarajan, Squire, and Suthiwart-Narueput (1996) discuss factors that should be considered to improve the economic analyses of projects. They suggest that project analysis and its supporting sector work should (1) establish a firm rationale for public involvement, (2) determine the counterfactual—what would happen with and without the project, taking into account the reactions of consumers and other producers, (3) identify the fiscal impact of the project on public funds; and (4) consider the possibility that loaned funds are fungible and that the real effect of the loan may derive not from the project identified, but from another project, chosen by the government and made possible by the additional funds. This paper examines these recommendations with respect to project evaluation in the health sector.

The World Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 47–71

The Rationale for Public-Sector Involvement in Health

The policy environment for projects today differs considerably from that of the late 1960s, when the basic ideas of contemporary project evaluation were formulated. At that time, economies in the developing world were highly distorted as a result of protectionist policies and government regulation or direct control of industry. It was assumed that governments would take a leading role in industrial projects, and the literature on project evaluation emerged as a way to help governments make socially profitable investment decisions. The question whether or not these activities should be in public hands was not an issue.

The world has changed significantly in the past thirty years. Countries have liberalized their economic policies and become more market oriented. The absolute level of distortions caused by taxes, trade barriers, and regulations has fallen, and many countries, including those that previously had centrally planned economies, have developed active and competitive private sectors. The premise that governments should carry out or rule on all investment projects is no longer accepted as given.

The techniques of project evaluation should adapt to this changed environment. Government investment, like any government intervention, should be justified in terms of the social benefit the project will have over and above that which would occur without public-sector involvement. For any investment opportunity, the focus of analysis should be on the *difference* between social and private benefits—not on the costs and expected returns to private goods themselves. The standard way to assess this relative benefit is to identify the market failures that characterize the private-sector equilibrium and to quantify the welfare loss from those failures. Priorities for investments should be based on the degree to which investments ameliorate welfare losses. Let me explain.

Policy formation in modern welfare economics usually begins by explaining how private markets allocate resources. In an ideal private market goods will be offered for sale to the extent that prices cover the cost to the seller of the last item sold. They will be purchased to the extent that their value to the buyer is at least as great as their price (and, hence, their cost). When prices, quantities supplied, and quantities demanded are mutually consistent, the market is said to be in equilibrium. *In the ideal market the subjective value to the buyer (private) is the same as the value to society (social), and in equilibrium both are equal to the cost to society of producing the commodity, thus ensuring that resources are allocated to their most efficient uses.*

Several circumstances interfere with the ability of private markets to operate with such efficiency, causing so-called market failures. These failures may occur, for example, if some goods and services are public goods, which cannot be withheld from persons who do not pay for them and which can be consumed by one person without reducing their availability to other consumers. The social value (total value) of such goods exceeds the private value of production, which is zero, because they will not be sold.

Related to public goods are goods that produce "externalities," benefits or costs to persons not party to the transaction and not considered by the producer or consumer when deciding how much of the item to sell or buy. The classic example of a negative externality is pollution, whereby a chemical factory does not figure into its costs the damage done by environmental dumping. It will therefore produce too much of its product, reducing the social value of the goods relative to their private value.

Monopolies by one or a few firms may also affect market equilibrium by permitting those firms to raise prices above marginal costs (the unit cost of additional production) and to restrict output, thus driving social costs above social values. Faulty information about the value or cost of products also affects equilibrium.

Associated with each of these market failures is a deviation of social and private value (or social value and social costs) and an associated social welfare loss that is, in principle, quantifiable in money terms.

If a "health project" is any investment in which the improvement of people's health is an important output, health projects comprise a very mixed assortment of activities. Some of these activities fall under the traditional jurisdiction of ministries of health; others, such as sanitation and safe water, may be directed by other ministries. They are mixed in another sense as well, because the components of health projects range from goods that are almost purely public to those that are almost entirely private, with services showing various degrees of market failure in between. Some principal areas of market failure in the health sector are noted below.

Infectious Diseases

Much has been made of the "epidemiological transition," the shift of the causes of mortality from infectious diseases to the noninfectious, chronic diseases typical of rich countries. Infectious diseases are still responsible, however, for a large proportion of deaths in poor countries, in particular, for deaths within poorer groups in those countries. They make a prima facie case for government intervention for three principal reasons.

First, they have distinct externalities. People with infectious diseases may not seek medical care quickly enough to avoid spreading the disease, and they may fail to complete a full course of treatment, which may lead to a resurgence of the illness, increased transmission, and increased risk of resistance to the drugs available for treatment. Similarly, for diseases for which transmission is decreased by the number of children immunized, an immunization program will confer an external benefit in addition to the benefit gained by the child immunized.

Second, some of the policy options available to fight infectious diseases are almost pure public goods; that is, no one can be excluded from using them, and use by one consumer does not preclude use by others. Goods of this kind will not be provided by the private sector. Control of disease-carrying pests (vector control) is one example. The collection and dissemination of health-sector informa-

Table 1. ***Female Adult Mortality Rates by Cause of Death and Income***
(percentage of females dying between the ages of 15 and 60)

Income quartile	*Infectious diseases*	*Noninfectious diseases*	*Injuries*
Richest	0.4	6.7	1.2
2	0.4	7.9	2.0
3	0.6	7.6	2.4
Poorest	1.4	8.9	2.7

Source: Adapted from Murray, Yang, and Qiao (1992).

tion, epidemiological surveillance, and laboratories to monitor safe food, water or drugs are others.

Third, infectious diseases disproportionately affect the poor. Table 1 shows the distribution of mortality by cause across different income groups of adult women in China. Although it is apparent that the poor suffer higher mortality from all causes, the rate at which they die relative to the nonpoor is very much higher for infectious than for noninfectious diseases (a factor of 3.5 as opposed to 1.3). If table 1 were extended to all age groups (in particular, to children), the relative effects of the kinds of diseases across income groups would be similar but the incidence of infectious diseases would rise relative to noninfectious diseases. Any reallocation from infectious disease control to noninfectious disease control in China would therefore hurt the poor most.

Uncertainty and Incomplete Information

A second set of market failures relates to the problems of uncertainty and incomplete information that plague the health sector. Although intervention is often suggested to correct these problems, it should be used with a great deal of caution. No market offers complete information about the goods and services it produces. If medical services are underused because consumers have incomplete information, an information campaign might be sufficient to correct the problem. If consumers, once informed, do not change their behavior, they may be expressing opinions about the value of the goods and services, rather than indicating a market failure regarding information.

This being said, imperfect information takes a few specific forms in the health field. First, for preventive health measures with no associated marketable product (the value of washing hands after defecation, for example, or of wearing long sleeves in the evening to protect against malaria), private mechanisms for delivering information may be inadequate. News sources sometimes cover these issues, but the value of news media as purveyors of information may depend on literacy or on the existence of a free press.

Second, the natural asymmetry of information in clinical health services may lead to "supplier-induced demand." Medical practitioners, who know more about health problems than their patients know, may overtreat for financial gain. Government policy can sometimes improve welfare in such markets, although

he appropriate policy depends upon the specific ways in which the market works. lthough some aspects of this problem have been modeled for the United States nd the United Kingdom, there are few good models for the developing world.

ealth Insurance and Efficiency Loss

Critical to the functioning of the health sector is the problem of catastrophic ss and the interaction of insurance markets with medical care markets (Griffin 990). Routine care is not necessarily difficult for consumers to handle out of ocket. It is the infrequent but financially devastating incidents that are of con-rn. Health expenditures in all countries are skewed toward a small fraction of he population that accounts for a large share of total health expenditure. Insur-nce markets may fail, and even cease to exist, when those who think they will eed care buy insurance, and those who think they will not, do not. The result-g higher costs of coverage will drive out still others who are relatively healthy, nd the entire market may unravel. This phenomenon is known as "adverse lection."

Problems of "moral hazard" may also occur with regard to insurance. The sured may be less careful in preventing an outcome that is covered by insur-nce or in minimizing the cost of getting service. The service provider, in turn, ay overcharge or overtreat insured clients in order to be reimbursed by the surer. These effects of moral hazard may then lead to suboptimal insurance overage, in that insurers may refuse to cover certain kinds of illnesses, treat-ents, or patients.

The absence of a well functioning insurance market means that large num-ers of people who would be willing to pay the actuarially fair rate to protect hemselves from catastrophic illnesses are prevented from doing so. The welfare ss associated with the absence of insurance markets is particularly relevant to he case of expensive procedures, for which consumer demand for coverage is ghest (Hammer and Berman 1995). In addition, the welfare gain per dollar ent to substitute for insurance will be highest for rare health problems, be-use these problems will have low expected costs against which to insure. The npredictable nature of the demand for health care, combined with the wide-pread absence of insurance is a key feature leading to large discrepancies be-ween the social and private benefits from care. The latent demand for medical surance (as opposed to medical care), and the efficiency loss induced by its bsence in developing countries is an especially neglected element in health eco-nomics. The recent Rand Corporation experiment in Indonesia is a notable ex-eption (Gertler and Molyneaux 1995).

The Alleviation of Poverty

Although the health sector is frequently called upon to help alleviate poverty, public intervention for this purpose should be treated with care. The kinds of

goods that are the best vehicles for redistribution though subsidized services are generally those that have very low (preferably negative) income elasticities; that is, they should be goods or services that poor people consume relatively more of than others consume. Although the health of the poor is worse than that of other groups, the poor do not generally demand more health care than others do. It is more often the case, in fact, that income elasticities for expenditures on health care are very high—usually greater than 1 and often close to 1.5. In other words relatively rich people spend a greater proportion of their income on health care than do the poor. Subsidizing services across the board, therefore, would transfer money to the wealthy. Van der Gaag (1995), citing examples from China, Côte d'Ivoire, Indonesia, Peru, and Tanzania, notes that many public health systems, although justified on the basis of ensuring equity, provide higher subsidies to the relatively affluent. Solon and others (1991), for the Philippines, show that people with high incomes receive much more of the marginal (additional) dollar spent on public health facilities than they pay of the marginal dollar collected in taxes.

As noted, however, the prevention of infectious disease usually helps the poor more than it helps others. The wide variety of possible health services thus has an equally wide variety of possible effects for different income groups. Before interventions in the health sector are designed with poverty alleviation as an objective, their effects on the ultimate beneficiaries need to be carefully calculated. Many kinds of health subsidies will have a perverse result.

Market Failure and Analysis

The fact that markets fail is not in itself a justification for intervention. It is also not a reason to ignore economic analysis. When markets work well, the standard prescription of laissez-faire policies is adequate. It is precisely when markets fail, and welfare losses occur, that they should be carefully examined to determine how they have failed and what measures will most effectively improve welfare. More attention needs to be given to the behavior of consumers, providers, and the markets for medical care and insurance.

Some of the analytical methods proposed for use in the health sector ignore these essential features. Despite recognizing that the public sector should not rely on the cost-effectiveness of medical procedures as an allocative criterion, the *World Development Report 1993* presents calculations of cost-effectiveness that are exclusively medical and that include none of the concerns discussed here (World Bank 1993, pp. 62, 117). Some readers have interpreted these numbers to mean that the higher the ratio of clinical benefits is to procedural costs, the higher will be the priority given to using public money for medical intervention.[2] Such calculations give no priority to infectious disease (or any other externality), no regard to the degree to which the private sector might substitute for the public, and no extra advantage to problems that disproportionately hurt the poor. With regard to risk and uncertainty, priorities based on the cost

effectiveness of medical procedures can get things backwards.[3] If the principal market failure in a particular context is the faulty insurance market, the items of highest priority for government intervention, from a welfare-improving point of view, should be those that are relatively expensive, holding possible health benefits constant. This preference is inversely related to the criterion of publicly funding the most cost-effective procedures (Hammer and Berman 1995).

Government Failures

Just as markets fail, so may government bureaucracies. Just as the behavior of private agents must be examined to judge how serious market failures may be, so must the behavior of public sector health service providers. The issue of monitoring quality and providing appropriate incentives within the public sector is discussed in more detail at the end of the paper.

Establishing the Counterfactual

The second goal advocated by Devarajan, Squire, and Suthiwart-Narueput is to assess what would happen with and without the project. Three aspects of the health sector make this goal particularly germane. First, health care in most developing countries is characterized by a substantial private sector functioning alongside a large public sector. Second, as a service, health care is largely a nontraded good. Third, health status, a primary output of the health sector, is difficult to value monetarily, leading to a need to account carefully and separately for health status in the net output of the project.

Taking the last first, agreement on a measure of the value of life is unlikely ever to be reached, and keeping separate account of the health effects of a project will thus be necessary. Such accounting requires knowing the actual *level* of consumption of services (rather than the *value* of consumption of tradable goods, as in standard analysis). As a nontraded good, consumption is equal to production, and, if a competing private sector exists, nontraded good production may "crowd out" (or possibly "crowd in") private production, leaving net changes in consumption as the focus of analysis.

Establishing the counterfactual therefore requires explicit modeling of demand for, and (nongovernmental) supply of, services. This underscores the value of identifying the market failure motivating the project. In this case, however, determining the behavior of the system will indicate the actual outcome of adding capacity to a market that has an active private sector, rather than simply providing a justification of the project. The standard literature on project evaluation takes nontraded goods into account by modifying the *prices* at which project outputs are valued (with the price capturing the net effect of project output on total market output). In the health sector, the reluctance to use prices on outputs such as lives saved means that the net contribution should be calculated

explicitly. The behavior of private-sector providers should also be analyzed to see if opportunities exist for improving services through regulation or subsidy that may be less expensive than direct public provision. Some elaboration of these points follows.

The Private Sector in Health

As table 2 shows, a large private sector is the rule in health care. With the likely underestimation of the use of traditional healers, the true size is larger still. This strongly suggests that the reaction of the private sector to public provision will be necessary to assess the net impact of the public intervention

Valuing Output

The difficulty in valuing outputs that entail extensions to life is one principal reason why health projects have been exempt from formal economic evaluation. Although I shall not contribute to the long and ultimately unsatisfying literature on undertaking this valuation, some judgment is needed to make informed decisions on public interventions in health. A few points relevant to practical project evaluation are noted here.

Valuation is simply a way of aggregating disparate inputs and outputs to get a single number as a measure of a project's profitability. Most of the time prices are the appropriate weights for this measure. The most visible problem in the health sector is the weight to put on life as opposed to money. Many other kinds of outputs within the health sector may be similarly difficult to compare—improving abilities to perform daily functions, for example, or relieving pain and discomfort associated with different diseases—and for which there is no market mechanism for valuation. Beyond an individual's appraisal of different health problems, the value to society of curing these problems for persons of different ages or functions is often debated in the literature, representing another dimension of aggregation. In addition, many aspects of the output of health systems

Table 2. *Public and Private Shares of Health Expenditures*
(percentage of total expenditures)

Region	*Public and foreign aid*	*Private*
Established market economies	61	39
Middle East	57	43
Former socialist economies	71	29
India	22	78
China	59	41
Other Asian economies and the Pacific Islands	39	61
Latin America and the Caribbean	50	50
Sub-Saharan Africa	53	47

Source: Govindaraj, Murray, and Chellaraj (1995).

are not specifically related to health. Time spent traveling (or away from work) to reach clinics, waiting time, the courtesy of service providers, and many other facets of a very personal service have been repeatedly shown to be important to consumers.

There is no correct solution to the problem. Any method of valuation chosen must be accepted as arbitrary, treated tentatively, and scrutinized seriously to ensure that policy conclusions are not sensitive to implicit assumptions. It is sometimes proposed, for example, that the present discounted value of a person's income stream be used as the value of life. This "human capital" method of valuation ignores the fact that retirees consider their own lives valuable, and there is no logical or ethical reason why such persons should be ignored in social calculations. This method can be effective, however, if an evaluation is made using lost human capital as a lower bound for the true cost of a disease and still shows a project to be worthwhile (Kim and Benton 1995).

An ideal measure of the value of different kinds of health outcomes would combine the personal preferences of patients (using their own assessment of discomfort, inconvenience, life prospects, and responsibilities) with a more objective appraisal of the medical effectiveness of care. Because this combination of knowledge—personal from the patient, technical from the provider—does not reside in any one person, it is fundamentally unobservable.

A method that most closely approximates this measure is the "quality adjusted life year" used in some industrial countries (Barnum 1995). This measure relies on extensive interviews asking respondents to trade off certain kinds of health problems against others. Even with this method, however, the number is an average and does not allow for individual variation in preferences.

Some of the other methods used have no way of incorporating preferences of patients. Measures such as "healthy life years gained" or "disability adjusted life years" make arbitrary judgments about the relative weights of different kinds of afflictions and the relative social weights of years of life lost at different ages. Anand and Hanson (1995) challenge the underlying logic and ethical judgments implicit in this measure.

Even if a defensible life value measure were obtained for a specific situation, generalizing it to contexts other than those captured in the measurement exercise would be difficult. A particular consideration is the degree of choice involved in the exposure to risks of death. One way of empirically estimating the value of life, for example, is to estimate wage differentials between safe and risky, but otherwise comparable, professions. Although the results of such studies are frequently interesting, they must be interpreted carefully. The subjects in the samples for the empirical work take risky jobs voluntarily. They may, therefore, be relative risk takers and not representative of the general public. Even if they differ little from others, however, there still remains the (ethical) concern that voluntarily taking risks is not the same as being exposed to risks without consent. Thus, deaths caused by motorcycle racing might be viewed differently from deaths from diseases caused by pollution (Viscusi 1992).

Although it has no solution, this problem cannot be avoided. It has sometimes been suggested that the difficulty in valuation can be circumvented by methods for which no value of life is needed. One such proposal is to use cost-effectiveness, which calculates the ratio of a given health impact from a medical intervention to its cost. Interventions with lower costs per health impact are then said to be preferred, and no explicit value of life is required. This strategy proves illusory in many of its proposed applications, however, in particular, in the choice between alternative treatment options in a clinical setting. It uses rate-of-return calculation to evaluate mutually exclusive options, a practice ruled out by the standard project evaluation literature (Hammer 1993a). An example of the way in which this method may yield unacceptable results is found in a paper by Sudre and others (1992), comparing different treatment options for malaria. Table 3 presents alternative program costs and the expected savings in lives for two different drugs. It is assumed that use of the drugs is mutually exclusive.

The authors note that if cost effectiveness is used as the criterion for deciding between the two drugs, chloroquine will be selected. If the larger number of lives saved by pyrimethamine-sulfadoxine is considered, however, the implicit value of a life will make the two equivalent ([value of life] × 1,382 − 1,812 = [value of life] × 1,723 − 2,622 or [value of life] = $2.38). They therefore conclude (p. 152) that "chloroquine would be the drug of choice only if the value of a death prevented were less than US$2.38 (but greater than US$1.31)." Not only is there an implicit value to life in the (supposedly value-free) use of cost effectiveness ratios, but it is absurdly precise and ridiculously low.

Because the problem cannot be avoided, the best advice is to be modest and to examine the logical consequences of alternative valuations. Health effects should be presented separately from other outputs (at whatever level of aggregation satisfies the policy analyst) to allow alternative estimates for the same value.

One way around the valuation of life issue is provided by the National Schistosomiasis Control project for Egypt (World Bank 1992a). For this project the rate of return was calculated under the assumption that the "switching value" that would make the project fail to pass a 10 percent rate-of-return test can be calculated and shown to be unreasonably low (Table 4). This method will not always give clear answers, however. Sometimes the value of life so obtained will be within a reasonable range for such a number. At the least, though, this calculation could give the policymaker something to discuss.

Table 3. *Costs and Effects of Alternative Treatments for Malaria*

Variable	*Chloroquine*	*Pyrimethamine-sulfadoxine*
Lives saved	1,382	1,723
Program cost (U.S. dollars)	1,812	2,622
Cost per life saved (U.S. dollars)	1.31	1.52

Source: Sudre and others (1992).

Table 4. *Rate of Return to a Schistosomiasis Control Program: Sensitivity Analysis*

Implicit value of a year of life (U.S. dollars)	Deaths averted each year	
	Number	
	4,600	2,300
	Rate of return	
800	40	17
600	28	12
400	18	n.a.

n.a. Not available.
Source: World Bank (1992a).

Determining Private Sector Behavior

As argued above, because medical care is a nontraded service, public production or provision (or financing) may displace services in the private sector. Any estimate of improved health status attributable to public expenditure should consequently account for the displacement of private services. The size of the effect is an empirical matter and should constitute a substantial part of the sector work leading up to the project. It can be derived from the overall market structure, which should have been a central focus of that analysis.

Substantial research has been conducted recently on the determinants of demand for health care in developing countries and the substitutability of public and private providers. Although the effort has been largely directed at determining the effect of public sector pricing on the use of health services (Akin and others 1985, 1986; Gertler and van der Gaag 1990), a growing number of studies have examined other aspects of demand likely to be affected by projects in health. A recent review by Alderman and Lavy (1996) examined the impact of location and quality of public health facilities on use. Table 5 reproduces some of the results.

The policy changes listed in table 5 are likely to be standard project components. Both can be counted on to decrease the numbers of persons who self-treat (do not visit a modern provider). When percent changes are weighted by the share of visits to each type of provider, however, they show that in Ghana 38 percent of new visits to public facilities in response to improved quality are attributable to reductions in visits to private facilities, as are 36 percent of visits in response to better access to public facilities. In Kenya fully 80 percent of the increased use of public facilities attributable to improved drug availability is accounted for by the drop in private facility use. If an evaluation were accurately to predict the increase in public facility use in response to the project but failed to account for the decrease in private sector use, the benefits (as some multiple of persons cured, say) would be overstated by a factor of five.

It is possible that the public sector provides better medical service than the private sector provides, and this quality differential should be examined (Ham-

Table 5. ***Effects of Public Facility Characteristics on Service Use***
(percent)

Policy simulation	*Self care*	*Public facilities*	*Private facilities*
		Percentage change in patient use	
Ghana			
Improve quality of care (infrastructure, materials, and staff)	–3.5	127.6	–19.5
Reduce distance to public facilities by 50 percent	–2.6	95.9	–14.9
Kenya			
Increase drug availability	-4.1	3.6	4.1
Reduce distance to public facilities by 20 percent	1.8	1.6	-1.8
		Share of market	
Ghana	51	14	35
Kenya	39	36	25

Source: Alderman and Lavy (1996), Ghana; Lavy and Germain (1994), Kenya; Mwabu, Ainsworth, and Nyamete (1993).

mer 1993b). It is also possible that improved access to free public facilities (in Kenya) is good for redistributive purposes. Redistribution depends on whether the average clinic user is poorer than the average taxpayer. If public clinics are disproportionately in urban areas and taxes come from agriculture, even the redistributive benefit is unlikely to be realized. In either case, to assess improvements in either health care or equity, the analysis underpinning the project should identify the market structure, the degree of substitutability and differences in the quality of public and private care, and the relevant characteristics of the beneficiaries (consumers).

Although the demand side of the market has been analyzed in some depth, the characteristics of the supply of services have been less well explored, and market analyses combining both supply and demand are rare. Gertler and Molyneaux (1995) provide one study that incorporates information from both the demand and supply sides of the market. They estimate the impact of public facility fees on private sector fees for Indonesia and find a close connection. Net demand changes are dependent on both prices.

Alderman and Gertler (1989) estimate the effect on demand for both public and private services in Pakistan of changing the public sector price of care. Although they have no data available to estimate the private-sector supply response, they explore possible net market effects by means of a sensitivity analysis. They find that the total effect of raising fees in health centers depends on the induced price rise in the private sector, because both prices are determinants of service use. In the context of project evaluation, the same kind of information might be used to examine the effect of making extra services available through the public sector (using quantities provided rather than fees charged).

Because direct information on the supply response of private providers is rare—the Indonesia study is unusual in that the private supply response was actually measured—experimentation with different values in a sensitivity analysis, as in the Pakistan study, is a possible solution.[4] In estimating the net effect of providing a competing service publicly, a simplifying assumption is that new public capacity enters the same market and has the same effect as new private capacity. If more detailed information suggests that the new public capacity has some other effect on the private sector, that effect should be included in sector work. New public capacity may reduce waiting times, for example, and it is time wasted waiting for free public services that generates the demand for private services. Whether or not the parameters (waiting time as determined by capacity, private sector demand as determined by waiting time) are known with certainty, estimates can be used to approximate the net effect of capacity. Alternatively, new facilities may decrease travel time, which was the source of private demand. Estimates of time savings and increased service use could be directly used in the project evaluation, combining information on demand as a function of distance with data on the geographical distribution of potential beneficiaries.

As these examples make clear, corrections for the impact of substitution with the private sector may be considerable. The degree of correction will be larger the larger the cross-price elasticity between public and private sectors, the larger the elasticity of supply of the private sector, and the smaller the overall elasticity of demand for services. Because many projects are long-lived (the expansion of clinic networks, the establishment of prevention programs), the relevant elasticity of supply is likely to be the long-run elasticity. Although this is more difficult to estimate accurately, it will probably be much larger than the short-run elasticity. In the short run, established private sector practitioners may not move from their current location or change the number of hours they work. With a longer time horizon, practitioners may decide to enter or leave a local market depending on how much the public sector draws potential clients. Similarly, potential professionals (university students) may choose to enter more profitable fields if the medical profession becomes less attractive.

Finally, the consumers elasticity of demand for services may differ greatly. Several studies have found that the price elasticity of demand for clinical care is higher for poor people than for others (Gertler and van der Gaag 1990). Projects targeted to reach the poor may therefore need to adjust less for displacement effects. Pritchett (1994) finds, with respect to contraception, however, that the number of children a family has is closely correlated with the number it desires. The demand for contraceptives is thus likely to be highly inelastic. That is because the cost of contraceptives is very small compared with the cost of having children. Contraceptive products are likely to be very elastic in supply (although methods requiring professional providers will share the supply characteristics of other medical services); if supply were inelastic as well, the prices of contraceptives would fluctuate widely. With elastic supply and inelastic demand, public programs subsidizing or providing family planning services can be expected to

have very little effect. Pritchett's empirical work confirms this expectation. Pitt, Rosenzweig, and Gibbons (1993) and Gertler and Molyneaux (1994) find similar results for Indonesia, using very different kinds of data.

Another method of determining the net outcome of projects is to estimate the effect of previous expenditures in the public sector on health outcomes. This strategy was followed in a World Bank analysis of Malaysia and repeated for several other countries (World Bank 1992b, 1995). The analysis, using a panel of regions within the countries, estimates the effect of different kinds of public expenditures (usually contrasting primary preventive services with subsidized curative, clinical care) on measures of health status, controlling for income. The results for Malaysia and the Philippines are reported in table 6.

For Malaysia robust results indicated that variations in traditional kinds of public health interventions (immunization and provision of safe water) were highly significant in explaining declines in infant mortality, but that public provision of clinical care services had no effect on health status. For the Philippines the results were less robust, but some specifications suggested the same conclusion. The best fitting specification, however, reproduced in the table, indicates something quite different. The presence of a significant interaction term between regional income and public health subsidies suggests that providing services in poor areas has an important effect on health status, whereas providing services in rich areas has no effect at all. The best explanation for this is that the (large) private sector substitutes closely with the public sector in higher income areas, resulting in a very elastic supply of private providers. In poor areas public

Table 6. *Determinants of Infant Mortality*
(coefficients from fixed effects panel estimation)

Variable	*Malaysia (14 states, 1986–89)*	*Philippines (13 regions, 1983–90)*
Income	-1.06	0.223
	(0.97)	(0.042)
Safe water	0.147	0.026
	(0.05)	(0.027)
Immunization: diptheria, pertussis, tetanus (DPT)	-0.113	0.018
	(0.04)	(0.013)
Publicly employed medical personnel per capita	1.03	
	(0.79)	
Public health expenditure		-0.404
		(0.113)
Public health expenditure × income		0.041
		(0.012)
Adjusted R^2	0.55[a]	0.988

Note: Standard errors are in parentheses.
a. Instrumental variables estimation, DPT as endogenous.
Source: Hammer, Nabi, and Cercone (1995); World Bank (1995).

rovision substantially increases access to health care because of a significantly ss-elastic private supply response.

For certain kinds of projects, such as some forms of vector control, sanitaon, or some kinds of health education and promotion activities, it is impossible charge for specific services, and there will thus be no private sector at all. For ch projects, no correction for the displacement of services need be made. For rvices in which private sectors compete with the public sector, the privatector counterfactual should be determined. In terms of substantive changes, is calculation is likely to raise the priority of population-based, public goods ojects, which have a substantial overlap with traditional public health internions, and to lower the priority of clinic-based, patient-initiated services, for hich the private sector can, and almost always does, exist.

This will not always be true, however, as the case of the Philippines illustrates. deed, consideration of the effect of projects on the private sector will not *necsarily* reduce the value of the public investment. The World Bank analysis of e health sector in Malaysia (World Bank 1992b), for example, suggests that e existence of a reliable public health service in that country has provided competition to the private sector that has effectively held down fees. This welfare provement from the public service could not have been estimated solely from e characteristics of the public program; it needed to be understood in the context of the industrial organization of the entire sector. The general lesson is that arket structures in poor countries differ substantially, and predicted effects of alth projects can go quite wrong if the preparatory sector work is lacking.

scal Impact, Fees, and Projects

he third point made by Devarajan, Squire, and Suthiwart-Narueput is that ce funds for investment come at a premium because of the distortionary ects of the taxes needed to collect them. Estimates in the literature indicate s that are on the order of 30 to 50 percent in industrial countries and that en greater for developing countries. In many poor countries, which may derdeveloped tax systems relying heavily on export taxes on agriculture, stributional effects of higher taxation can make this cost even higher. This n leads to three main conclusions relevant to health sector projects. First, many projects are accepted using conventional project evaluation methods. cond, opportunities for recovering costs in the project should be explored. lternative, perhaps cheaper, methods should be examined for correcting e market failure that initially justified the project. Regulations, partial subsies, or even taxation may be as effective as project funding in improving the arket outcome.

As applied to the health sector, the first conclusion suggests that public expenditures should correct only market failures for which the welfare costs are at east as high as the damage caused by the taxes needed to finance them. Because

formal project analysis has not been applied in health and so cannot have led to a bias in favor of public-sector delivery, ignoring the cost of public funds has probably led to a similar, if informal, bias.

The second conclusion, that options for cost recovery should be explored lands squarely in the middle of a long-standing controversy in the health field (Creese 1991; Griffin and Shaw 1995). A few points from the analysis in Devarajan, Squire, and Suthiwart-Narueput might clarify this issue. First, many of the services offered in health care are private goods, whether or not they are delivered by the private sector. Fees can be charged for these services and nonpayers excluded from them, even if charges are not currently levied. If there is a premium on public funds (or a budget constraint for the health ministry the decisions about which projects to support and how much to charge should be made jointly.

Whether and how much to charge depends on balancing two opposing concerns (Hammer 1993b). Because charging fees will reduce the drain on the government budget for a given project (or allow a fixed budget to be stretched further), cost recovery will translate into a higher priority for any particular project. Raising fees will, however, reduce demand for the publicly provided services. The question to ask is what the consequences are likely to be of failing to get that care. How many consumers will be dissuaded by higher fees from seeking publicly provided care, and what difference will it make to their health?

The effect on individual health depends on several factors. First, do people stop seeking treatment altogether, or do they switch to care provided by the private sector? Are the health conditions for which they stop seeking care likely to be serious or not? Second, what is the relative effectiveness of treatment in the public and private sectors? To put it starkly, if as a result of higher fees people are staying away from clinics for treatment of muscle aches and skin rashes (a large component of demand for local hospital services in Indonesia) or are buying the same over-the-counter treatments that they would get from the public facility, that is one thing. If they are sitting home, infecting others, and dying of tuberculosis, that is quite another

Who benefits from fee increases, and how, also depends on whether the money collected is retained by the clinics and used to improve the quality of services. Jimenez (1987) shows the conditions under which the improvement in quality can outweigh the financial burden of the fees. Litvack and Bodart (1993) demonstrate just such an effect in Cameroon and argue that poor people in particular benefit in terms of overall access to services from the combined effects of fees and improvements in quality.

Fees are least likely to harm the health status of consumers when (1) the demand for care at public facilities is inelastic, that is, when higher fees do not dissuade clients from using public clinics; *or* (2) demand at public facilities is elastic but clients continue to use the facilities for more serious conditions and stop using them for minor ailments; *or* (3) demand at public facilities is elastic but private facilities are close substitutes, that is, the cross-price elasticity of

emand is high and private supply is also elastic, meaning that clients stop using ublic facilities but still receive care at private clinics; *and* (4) private care is ffective. If the private sector is characterized by modern providers, such as ongovernmental organizations (NGOs) or public providers in their off hours (as s legal and standard practice in Indonesia), a shift in use from public to private ervice does no harm to health status. If the private sector consists of traditional ealers with no particular skills (not true of all traditional healers), increasing emand for their services is harmful.

To judge the effect of fees, therefore, it is necessary to know a fair amount bout the demand for services and the nature of the supply of private services. s mentioned above, market characteristics such as demand elasticities (some-imes with cross-price effects) are known from research but vary substantially rom country to country. They cannot be confidently inferred from one country o another and should be investigated in the specific context of the sector work nder way. Some market characteristics need to be examined more thoroughly, uch as symptom-specific demands for services, which can help explain whether eople stop using life-saving care. Note that information about the cost-ffectiveness of medical treatments is *not* part of the essential information needed o determine which services should be provided at subsidized rates in the public ector. Cost-effectiveness shows up only with regard to the difference in effec-veness between public and private care.

Third, the high cost of public funds implies that more effort should be given o looking for policies other than subsidized provision or financing to correct arket failures. Monitoring and regulating a private sector may be such an ption for ensuring high standards of care at a lower cost to the government. The same might be true for private insurance. Similarly, if imperfect informa-on is a key element in the health market, providing information concerning the uality and effectiveness of private providers can be an important public role van der Gaag 1995).

The information needed to choose between regulating and providing services s unfortunately, usually lacking. Similarly, the functioning of insurance mar-ets in the health sector is not sufficiently understood to be able to prescribe onfidently the appropriate regulatory framework. Recent experiments in the rovision of insurance should shed light on this issue (Griffin and Shaw 1995).

Fungibility and Other Issues of Public Servant Behavior

Devarajan, Squire, and Suthiwart-Narueput note that the true effect of project unds may have little to do with the specifics of the project being evaluated. The overnment may have intended to start the project anyway, and the extra money imply allows it to finance another project that it may have considered mar-ginal. Donors may be unaware of the project they are actually funding, much ess be able to evaluate it. Much of the force of this argument comes from the

significantly larger scope for reallocations between, rather than within, secto but it also has relevance to possible reallocations within health ministries. Sin the Alma Ata conference in 1977, the international public health commun has stressed the need to shift resources toward basic primary care. As a resu much of the public health funding available from donors has been directed t ward primary care. When governments' actual allocations are examined, ho ever, evaluators find that large portions of the budgets are directed toward se vices that do not conform to the primary care model.

Source of 'Government Failure'

The fact that actual allocations differ so substantially from the internatior paradigm suggests that the ministries' decisions are determined by internal fa tors, such as political pressure from providers or affluent consumers. It is al entirely possible that governments count on project financing from donors fund the basic services (immunizations, rural care), leaving the ministries free satisfy pressures to provide or subsidize urban, tertiary services.

A recent paper by Feyzioglu, Swaroop, and Zhu (1996) analyzes the effect foreign aid on public expenditure patterns. Within the health sector, the autho find that although foreign aid earmarked for the health sector has reduced i fant mortality in recipient countries, the governments' own resources spent health have not. This finding suggests that governments' allocations comp sate for the preferences of donors, and it reinforces the point made by Devaraj Squire, and Suthiwart-Narueput that projects should be evaluated in the co text of the overall sectoral strategy or by reviews of public expenditures acr the board. To the extent that money is fungible within ministries, this adv seems warranted.

Incentives

Critical to effective project analysis is an understanding of the goals, ince tives, and constraints of governments and their workers. Behavior is at the hea of "government failure," fungibility, and the decision to provide, rather th regulate, services. The incentives facing senior policymakers, as well as the behavior, must be known to understand the fungibility of resources. To unde stand the true impact of a project, the same information must be gathered ab the civil servants responsible for project implementation. To decide betwe providing and regulating services, the incentives to private providers—as well a to civil servants—their probable behavior, and the ability of policymakers t influence those incentives must be considered.

Although considerable intellectual effort has gone into defining the right price by which to value outputs of projects, the inputs and outputs themselves are largely treated as given. The guideliness drawn up by UNIDO (United Nations Industrial Development Organization) were one of the original standard texts in

he field; they state that project evaluation is divided into ten steps. The first is to "ascertain the 'net output' of the project and split it into adding to supply and saving resources" (UNIDO 1972, p. 50); the next nine deal with shadow pricing (imputed valuations) and the like. In addition, much of the discussion of the first step concerns splitting up the outputs; almost nothing is written about "ascertaining the net output."

It is naive to assume that project inputs will achieve their intended results irrespective of the incentive structure facing those responsible for them. These incentives may be specific to the individuals actually constructing the project and running the enterprise (if it is kept in public hands) or may affect private agents (such as farmers in an irrigation project) during the period when project benefits are generated. The former category raises issues of ownership at the higher levels of supervision and issues of civil service remuneration or contracting procedures at the lower levels. The latter, which has been the subject of numerous analyses, depends on the policy framework within which the project operates. In a recent analysis of incentives, Pritchett (1996) finds that the discrepancy between the value of capital as determined by accumulated costs of investments and its value as determined by contribution to output (marginal productivity) varies enormously across countries. This discrepancy is attributed to the economic environment in which the investments have been made and argues strongly against using simple input output relationships that are independent of the incentives faced by people working with invested capital.

The incentive structure is especially important for evaluations of health sector projects, because health care is a service. Its value therefore depends upon maintaining incentives to service providers for sustained good performance. Just because a health clinic is built does not mean the providers will show up for work. And if they do come to work, there is no guarantee they will devote themselves to the care of their patients. The actual output of the investment depends on policies concerning pay and other incentives for good performance in public employment. Just as incentives facing people in the private sector should be examined for evidence of market failures, so should incentives in the public sector be examined for evidence of government failures.

Recent analyses of public health systems point to serious problems in this regard. One indication of these problems is the common bypassing of local public health facilities for private (or higher-level public) facilities even when the public service is free (Kloos 1990; Korte and others 1992). Among the reasons given for this are lack of concern shown by the public provider, social distance between the medical practitioner and his or her clients (exacerbated by assigning doctors to ethnic areas different from their own), and other aspects of the behavior and degree of commitment of the civil service doctor.

Lewis, Sulvetta, and LaForgia (1991, 1996) point to profound problems in the technical efficiency of public hospitals and clinics. In one study in the Dominican Republic, the proportion of expenditures actually reaching patients as services is estimated to be as low as 12 percent. Once again, the incentive struc-

ture, in this case for hospital administrators, is central to the problem. If th
financial viability of the enterprise has no impact on pay and promotion, th
quality of management is likely to suffer.

In sector work on Indonesia (World Bank 1994), a particular dilemma wa
identified. Regional variation in epidemiological conditions and the variety o
tasks expected of public health employees argued for increased local discretio
in the allocation of resources. The incentive system in place, however, whic
allowed doctors to maintain private practices in public facilities (in the afte
noon when the public facilities were closed), raised the possibility that this di
cretion would be used perversely, leading to heavier reliance on clinic-base
activities to the detriment of outreach and population-based public health con
cerns. Reforms were recommended to compensate public providers who bette
served public priorities.

It is widely believed that a consequence of fee-for-service private care is th
tendency of doctors to overtreat so that they gain more income from more ex
tensive service. A fair question is whether the same doctor on salary will undertrea
or badly treat if there are no financial consequences for failing to satisfy custom
ers. Both problems can, in principle, be addressed by careful monitoring an
appropriate sanctions. As important as the issue is, however, little is know
about the relative abilities of governments to manage public systems or regulat
private systems.

Conclusion

This article has tried to show how the issues raised by Devarajan, Squire, an
Suthiwart-Narueput might be applied to project evaluation in health. Althoug
the health sector is characterized by several market failures that may justify
public-sector involvement, the existence of imperfectly competitive markets
should not be used to ward off economists and to justify simply any interven
tion. Specific market failures should be identified, and the analyses identifying
them should give guidance on ways to correct for them.

The inadequacy of insurance markets is one characteristic of the health sec
tor. This market failure should lead to an analysis of, and attempts to measure
the value of the reduced insecurity that a project might offer to beneficiaries
Also characteristic of the health sector are services that either are pure public
goods (pest control) or have distinct external effects (infectious disease control)
In addition, health projects are often promoted to alleviate poverty. Different
kinds of services, however, have very different distributional characteristics. Basic
sanitation, hygiene, and even the education of girls from poor families may have
a greater impact on the health of the poor than will general subsidies to clinical
services. It should be noted that none of these three areas—risk, external effects,
and poverty alleviation—are handled by current applications of cost-effective-
ness analysis.

With regard to establishing the counterfactual, the trick is to give a full ac-)unt of what will happen with the project *taking into account the reaction of ther actors in the health system*: the consumers, private providers (including ·aditional healers), NGOs, and insurers (including informal credit or private trans- ›rs). Because health is largely a nontraded good, the participation of the gov- rnment will affect the overall consumption of health services. Being aware of ich features of the medical service market as demand elasticities (with respect) prices as well as to project characteristics such as service location or quality) nd supply elasticities of private providers (particularly with regard to price or ie existence of a competitive public sector) is critical to being able to establish is reaction.

The third element, determining the fiscal impact of the project, suggests that ie premium on public money should lead to a search for alternatives to subsi- ized provision. In health the setting of fees for clinical services is the most nportant area in which this issue arises. This paper argues that higher subsidies hould be directed toward services that have higher social returns relative to ivate returns and, of those that warrant subsidy, to those that have more elas- c demand. It is important to know when raising prices will be counterproduc- ve (for example, for services with high price elasticities) and when money can e saved for other high-priority needs by charging fees (for example, for services ith low price elasticities).

The fourth component of good project analysis, acknowledging the fungibil- y of project resources, relates to the much broader issue of understanding the ehavior (motives and incentives) of public servants. The motives of high-level ublic servants are likely to be an issue in the health sector if donor funds ear- narked for primary health care free up domestic funds for services on which hose donors would place low priority. More fundamental, however, is under- tanding the incentives that will encourage civil servants or government con- ractors to provide high quality, responsible care. Virtually all projects assume at project inputs are used appropriately and that the output is known (at least terms of numbers of patients seen at some assumed level of quality). This ssumption is not valid for analyzing services, and the incentive structure of ervice providers should be a far more important topic for analysis and research han it has been.

Much of the analysis relevant to projects should be done before the project valuation stage. Indeed, given the issues of fungibility and incentives, the est form of intervention by donors may not be through traditional projects t all but, rather, through general loans conditioned on overall sector strat- gy and reform. If a standard project is proposed, however, considerable nformation from supporting sector work will be needed before evaluation. In particular, if clinical or other services that require public participation are part of the project, the supply and demand for substitute services will need to be known. Just as shadow (imputed) exchange or wage rates for project evaluation should be derived from supporting economic work done before

the specific project, so should the market structure of health be part of th
background investigation.

Adequate market analyses are rare in developing countries. What, then, shoul
project evaluation look like before such analyses become available? First, evalu
ators should be serious about collecting information. In the interim, project
may be designed to focus on the principal market failures that such research i
likely to uncover. Evaluations of health activities that are pure public goods o
that address genuine externalities will be little affected by more detailed mark
analysis and so may proceed. For projects in which poverty alleviation is cen
tral, it should be established that the project beneficiaries are, in fact, the poor
Health care as a redistributive device should be used with caution, unless [illegible]
geographic placement or other means, the poor can be encouraged to take ad
vantage of the health services and others discouraged. For other projects, w [illegible]
large clinical components, more scrutiny is needed. Without knowing much abo
the market the government will enter or the incentive structure for public pro
viders, it is difficult to know what health effects (or even service use) will res
from a project. Estimates of the elasticity of demand for services specific to the
country and guesses concerning the elasticity of competitive supply should be
incorporated as weights on predicted project outputs. Identifying these param
eters should be the focus of future analyses of the sector, and the absence o
information on these parameters should encourage caution on the part of proje
evaluators.

Notes

Jeffrey Hammer is principal economist with the Public Economics Division of the [illegible]
Research Department at the World Bank. The author thanks Shanta Devarajan, [illegible]
Golladay, Maureen Lewis, Sandy Lieberman, Phil Musgrove, Lant Pritchett, [illegible]
and Sethaput Suthiwart-Narueput for discussions on the issues considered in this [illegible]
and for comments on an earlier draft.

1. "Using cost-effectiveness to select health interventions for public financing does
necessarily mean spending the most resources where the burden of disease is g[illegible]
Instead, it means concentrating on the interventions that offer the greatest possible [illegible]
health per public dollar spent. The relevant comparison is usually not with a situat
which nothing is done but with the situation created by privately financed health interv
tions" (World Bank 1993, p. 65).

2. Bobadilla and Saxenian (1993, p. 10), for example, precede their reference [illegible]
medical intervention numbers by stating: "*That is why the first step in designing* [illegible]
essential health package is to determine the cost effectiveness of a health intervention [illegible]
net gain in health compared with doing nothing, divided by the cost."

3. The contribution a particular health problem makes to the overall burden of dis[illegible]
has also been suggested as a criterion for setting priorities for public intervention (M[illegible]
and Lopez 1994). This is also incorrect for contexts in which the main market failure [illegible]
insurance. The health conditions that will be most seriously underfinanced (in welfare
terms) in the absence of an insurance market are the rare ones, rather than the common
ones. With a few exceptions, the total impact of a problem is irrelevant to decisionmaking,
the pertinent criterion being the marginal impact a policy will have in correcting the pr[illegible]
lem. The exceptions are pure research in which neither the potential cost nor the like[illegible]

———. 1993b. "Prices and Protocols in Public Health Care." Policy Research Workir Paper 1131. World Bank, Population and Human Resources Department, Washingtoı D.C. Processed.

Hammer, Jeffrey S., Ijaz Nabi, and James Cercone. 1995. "Distributional Effects of Soci Sector Expenditures in Malaysia, 1974 to 1989." In Dominique van de Walle and Kin berly Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore, Md The Johns Hopkins University Press.

Hammer, Jeffrey S., and Peter Berman. 1995. "Ends and Means in Public Health Policy Developing Countries." In Peter Berman, ed., *Health Sector Reform in Develop Countries*. Cambridge, Mass.: Harvard University Press.

Jimenez, Emanuel. 1987. *Pricing Policy in the Social Sectors: Cost Recovery for Educati and Health in Developing Countries*. Baltimore, Md.: The Johns Hopkins Univers Press.

Kim, Aeh-Yung, and Bruce Benton. 1995. *Cost-Benefit Analysis of the Onchocerciasis C trol Program*. World Bank Technical Paper 282. Washington, D.C.

Kloos, Helmut. 1990. "Utilization of Selected Hospitals, Health Centers, and Health St tions in Central, Southern, and Western Ethiopia." *Social Science and Medi* 31(2):101–14.

Korte, R., H. Richter, F. Merkle, and H. Gorgen. 1992. "Financing Health Services in S Saharan Africa: Options for Decision Makers during Adjustment." *Social Science Medicine* 34(1):1–9.

Lavy, Victor, and Jean-Marc Germain. 1994. *Quality and Cost in Health Care Choice Developing Countries*. Living Standards Measurement Study Working Paper 105. W ington, D.C.: World Bank.

Lewis, Maureen, Margaret Sulvetta, and Gerard LaForgia. 1991. "Productivity and Q ity of Public Hospital Medical Staff: A Dominican Case Study." *International Jo of Health Planning and Management* 6(4):287–99.

———. 1996. "Measuring Public Hospital Costs: Empirical Evidence from the Domi Republic." *Social Science and Medicine* 43(2):221–34.

Litvack, Jennie, and Claude Bodart. 1993. "User Fees and Improved Quality of He Care Equals Improved Access: Results of a Field Experiment in Cameroon." S *Science and Medicine* 37(3):369–89.

Murray, Christopher J. L., and Alan D. Lopez. 1994. *Global Comparative Assessme the Health Sector: Disease Burden, Expenditures, and Intervention Packages*. Ge World Health Organization.

Murray, Christopher J. L., Gonghuan Yang, and Xinjian Qiao. 1992. "Adult Mo Levels, Patterns, and Causes." In Richard G. A. Feachem, Tord Kjellstrom, Christo J. L. Murray, Mead Over, and Margaret A. Phillips, eds., *The Health of Adults Developing World*. New York: Oxford University Press.

Mwabu, Germano, Martha Ainsworth, and Andrew Nyamete. 1993. "Quality of Me Care and Choice of Medical Treatment in Kenya: An Empirical Analysis." *Jour Human Resources* 28(4):838–62.

Pitt, Mark, Mark Rosenzweig, and Donna Gibbons. 1993. "The Determinants and C quences of the Placement of Government Programs in Indonesia." *The World Economic Review* 7(3):319–48.

Pritchett, Lant. 1994. "Desired Family Size and the Impact of Population Policies." *Popu lation and Development Review* 20(1):1–55.

———. 1996. "Minding the P's and Q's: The Cost of Investment and the Value of Ca tal." World Bank, Policy Research Department, Washington, D.C. Processed.

Solon, Orville, and others. 1991. "Health Sector Financing in the Philippines." Research Triangle Institute and University of the Philippines School of Economics, Manila. Pro cessed.

Sudre, Philippe, Joel Breman, Deborah McFarland, and Jeffrey Koplan. 1992. "Treatment of Chloroquine-Resistant Malaria in African Children: A Cost-Effectiveness Analysis." *International Journal of Epidemiology* 21(1):146–54.

UNIDO (United Nations Industrial Development Organization). 1972. *Guidelines for Project Evaluation.* New York.

van der Gaag, Jacques. 1995. *Private and Public Initiatives: Working Together for Health and Education.* Directions in Development Series. Washington, D.C.: World Bank.

Viscusi, W. Kip. 1992. *Fatal Tradeoffs: Public and Private Responsibilities for Risk.* New York: Oxford University Press.

World Bank. 1992a. "Arab Republic of Egypt: National Schistosomiasis Control Project." Staff Appraisal Report 10513-EGT. Washington, D.C. Processed.

1992b. "Malaysia: Fiscal Reform for Stable Growth." East Asia Department I, Washington, D.C. Processed.

1993. *World Development Report 1993: Investing in Health.* New York: Oxford University Press.

1994. "Indonesia's Health Work Force: Issues and Options." East Asia Department 3. Washington, D.C. Processed.

1995. "Philippines: Public Expenditure Review." East Asia Department I, Washington, D.C. Processed.

NEW FRONTIERS IN PROJECT EVALUATION? A COMMENT ON DEVARAJAN, SQUIRE, AND SUTHIWART-NARUEPUT

Arnold C. Harberger

Readers might infer from the paper by Devarajan, Squire, and Suthiwart-Narueput and from Hammer's applications of that methodology to health projects, that the authors are proposing fundamental modifications of ...ndard techniques and of received theory of cost benefit analysis (other-... known as applied welfare economics and in some uses as social or eco-... project evaluation). Such an inference is not warranted, however. So far ... see, nearly everything the authors propose fits quite easily within the ...ed corpus of applied welfare economics. The steps that they advocate are ...cations not of standard cost benefit analysis, but of habits that have de-...ed over the years and decades both in the World Bank and quite generally ... practitioners of economic project evaluation.

...mmer neatly summarizes Devarajan, Squire, and Suthiwart-Narueput's main ...ption: there should be a firm rationale for public involvement if a project is to ... in the public sector, the project should be compared with a clear ...factual", the fiscal impact of the project should be clearly identified and ... be assigned a properly estimated cost, and the issue of fungibility of funds ...be clearly addressed in weighing the economic consequences of project loans.

...blishing the Counterfactual

...aps just as professors in introductory economics courses harp on the dif-... between "average" and "marginal" costs, revenues, productivity, and ...rofessors of project evaluation try to hammer home the dichotomy of ...ersus "without" a project. Ideally, we build a running scenario of the ...without" the project, a moving picture going ahead ten or twenty years, ...ding on the project's economic life. Into this scenario we then insert the ...ct, duly noting all the relevant differences between the two scenarios. These ...nces are the basic inputs into project evaluation. Usually we start with the

...Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 73–79
The International Bank for Reconstruction and Development / THE WORLD BANK

project's own purchases and sales or other flows of direct costs and benefits. But we really should go on to assess differences that arise outside the project. This is what we are doing when we use, say, the economic opportunity cost of foreign exchange, or of capital, in project evaluation. These opportunity costs capture the fact that the use of foreign exchange in the project displaces other imports. The tariff revenue lost on these imports is captured in the economic opportunity cost of foreign exchange. If the project pays tariffs on some of its own imports, *those* tariffs are captured as part of the difference between the project's financial benefits and its economic benefits.

Economists as a profession are pretty good at capturing the external effects that are summarized in the opportunity costs of capital and foreign exchange because these are visualized as being generated when any new demand is inserted into the corresponding (foreign exchange or capital) market. Economists are not nearly so good at capturing types of displacement that are unique to each project. Not all evaluators even look for them; others may look in a perfunctory way and turn up some, but not all, of the relevant externalities. I know of no one, even among those who have made the most serious efforts, who would state flatly that no important externality has escaped his or her net (in a real-world project evaluation).

Economists know what they should do in principle, and they recognize this when they give lectures and ask or answer examination questions. But when they are out in the field, evaluating real projects with limited time and resources they practice far less than they preach.

Issues of Public-Sector Involvement

Traditional project evaluation and traditional applied welfare economics have been color-blind about whether a project is located in the public or the private sector. All they are interested in is the profile of the project including all relev externalities. If profiles A and B are the same, the projects get the same evalu tions, independent of the sector in which they are situated.

I do not believe that Devarajan, Squire, and Suthiwart-Narueput would agree with this statement, but they might say several other things. First, many public-sector projects have been evaluated without properly tracking t counterfactual. If a public-sector textile project displaces a certain amount private-sector textile capacity, then such things as the taxes that would ha been paid by the private sector but are not paid with the public-sector proje must be counted as costs.

Second, they may impute to the public sector a lower degree of efficiency a slower adaptation to economic change—or both—than the private sector. T assumption here is that private firms will adopt cost-reducing technologies mo quickly and will abandon nonviable lines of production more rapidly than w public sector firms. Economists used to find this hard to say, but they now

derstand that the public sector's sluggishness in these respects arises out of real political pressures. Recognition of this fact has been one of the factors behind the recent worldwide wave of privatizations. The current response to pressures that prevent state-owned firms from being economically efficient is to privatize those firms and thus get the government out from under.

Thus if the project is in the public sector and can be expected to supply goods to the public at a higher real cost, on average, than would a private producer, that should be enough to condemn the project to the discard pile, for one of the most fundamental precepts of project evaluation is that one should not ascribe to a given project a benefit that is greater than the alternative cost of providing that (or an equivalent) benefit by another route.

I believe that in these and similar cases, old-fashioned project evaluators would have followed a path very similar to the one recommended by Devarajan, Squire, and Suthiwart-Narueput. The nuance of differences that might exist (and I am far from sure that it does) could be that Devarajan, Squire, and Suthiwart-Narueput might really prefer the private-sector project to the identical public-sector project, which comes close to implying that they might opt for the private-sector project even if its true benefits were somewhat less than the public-sector project.

A Shadow Price for Fiscal Funds

On this point I have a confession to make. For some three decades I have ardently defended the convention, in cost-benefit analysis, of assuming that the marginal source of project funds was (government) borrowing in the capital market. I did so (and still do) because this is where marginal increments of money typically come from (as governments increase debt to cover deficits, both foreseen and unexpected) and go to (when either foreseen or unexpected surpluses appear). This convention is useful because the capital market (the banking system in some developing countries) is the actual source to which governments turn. It is also useful in that the ultimate sources of funds are determined by elasticities of supply and demand in the capital market. These are likely to be far less capricious than the weights that would apply to successive tax changes, which tend to be very different in their impact from one tax law (eliminating investment credits) to another (an alcoholic beverages tax or an increase in a value added tax) to yet another (a reduction in income tax rates and a partial integration with the corporate income tax).

I thought I had made a very strong case for a capital market "convention." And I felt little pressure to make a stronger one, because nearly every framework for project evaluation uses the capital market convention anyway. So there was no vocal group out there clamoring to be convinced.

My complacency was interrupted, about a year ago, when I was writing a sort of current status report on economic project evaluation. One of the topics I

was asked specifically to address was that of shadow prices for fiscal funds. My first reaction to the idea of such a shadow price was negative, because I thought that it meant abandoning the widely accepted convention of sourcing in the capital market. But further reflection revealed to me that it meant no such thing. Indeed, it was not only fully compatible with that convention, but in fact was arguably a natural complement to it.

First, consider a public-sector project that just pays for itself in budgetary funds, discounted at the economic opportunity cost of capital. For such a project one can either use a shadow price for fiscal funds or not. If fiscal costs equal fiscal benefits, then fiscal costs multiplied by $(1 + \lambda)$ equal fiscal benefits multiplied by $(1 + \lambda)$. [$(1 + \lambda)$ is the shadow price of a dollar of fiscal funds.]

Now consider a project that does not generate enough fiscal revenue to cover its costs. One option would be to assume, following the convention, that the funds (say, $10 million) initially came from the capital market and that project revenues were sufficient to pay back, say, $6 million (in present value terms). What should be done about the remaining $4 million? My old assumption seemed to imply that this $4 million of present value gets incorporated into the debt that never will be repaid. Seen in this light, the assumption does not seem too wise.

The next step was to assume that at some point during or after the life of the project, the government would use fiscal revenues to pay off the project's remaining debt. Raising those revenues would carry an excess burden because all real-world taxes are distortionary. It is difficult to measure the excess burden because the pattern of tax rates that will be used to cover the project's deficit cannot be predicted. This particular dilemma does not disappear when one moves from a "fiscal finance convention" (rejected by me and by most) to a "shadow price of fiscal funds" (proposed by Devarajan, Squire, and Suthiwart-Narueput and accepted at least by me). My suggestion at this point is that one should work with the lower end of the plausible values of λ (the excess cost per dollar of fiscal funds). That would mean leaning toward the 0.17 end of Devarajan, Squire, and Suthiwart-Narueput's reported range of 0.17 to 0.56 (for the United States).

In any case, if one assumes that the project's fiscal deficit is covered by its funds raised by taxes, one might think of applying the factor $(1 + \lambda)$ just to fiscal deficit. But once again, simple accounting identities come to the rescue of economists. One way to apply this factor to the fiscal deficit of the project is to apply it to all the fiscal outflows and all the fiscal inflows, that is, to treat it as a shadow price of fiscal funds generally. I take this to mean all the cash outlays of the project that are done with government money (own or borrowed) together with all the inflows from the project into government hands.

My final conclusion is that I am not at all troubled, indeed I am pleased with the idea of using a shadow price of, say, 1.20 or 1.25 for all fiscal flows on a project. I believe this order of magnitude would be on the low side (as I think it should be) not only for the United States, but for nearly every other country.

On the Fungibility of Funds

Standard project evaluation does not deal with the fungibility of funds because it focuses on the project rather than on the means of financing. Put somewhat differently, the capital-market-sourcing convention says that the money spent on the project could always simply be dumped back into the capital market, where, by definition, its true economic return (counting all externalities) would be the opportunity cost of capital. That is why the economic opportunity cost is used as the discount rate. (There is some literature that advocates treating "soft loans" differently from regular sources of funds, but I believe that most serious thinking on this subject comes to the conclusion that all projects should pass the test, using the opportunity cost of funds as the discount rate, no matter how much soft money happens to be put at the project's disposal.)

I believe that these observations imply an implicit recognition of the fungibility of funds, but it is from a different angle. Devarajan, Squire, and Suthiwart-Narueput look at the problem from the standpoint of the lending agency, and when push comes to shove, they seem to be as interested in evaluating the loan as in evaluating the project. This is perfectly sensible and straightforward. In fact life becomes very simple once one recognizes that two different evaluations are involved.

If one focuses on evaluating the loan, one can bypass the project that it ostensibly finances and try to track down the project that it really ends up financing. Trackers will find the jungle very murky here. Rarely will it be possible to identify the specific projects that the loan will finance.

What is the solution? My opinion has always been that a good alternative for international lending agencies as well as foreign aid agencies is to insist that not only must the project be evaluated by sensible criteria, but also that the host country use similar criteria to evaluate its other projects.

This may be close to what Devarajan, Squire, and Suthiwart-Narueput presume, but I think there is some element of difference. I would emphasize the importance of a country's developing (and institutionalizing) its own project evaluation capability. I have participated in a few exercises in this direction and I assure readers that it is a difficult row to hoe. It is critical to have technically competent (and morally upright) people in charge of such an effort. One does not tend to have that at the beginning. So the row turns out to be a long one, with the countries and donor agencies planting the seeds and nurturing the seedlings as they grow.

This sort of work is a prior step to what Devarajan, Squire, and Suthiwart-Narueput recommend for sectoral public expenditure reviews. But certainly we are all on the same side of the fence when it comes to prescribing how to deal with the problem of fungibility. My position is fully compatible with their call for periodic sectorwide reviews of the results of past projects, the likely need for new ones, and public expenditure allocations in general. A great deal of attention has been focused on limiting the size of government and enhancing its effi-

ciency, but I suspect that old-fashioned project evaluation theory, if well applied, would not only lead to a similar conclusion, but to one that was based on considerations of economic efficiency.

One way to limit the government's reach would be to impose, in evaluations of public projects and programs, a penalty on the use of funds by the government. This is where we come out in the end, but the penalty of λ is motivated in both our cases by standard efficiency considerations—the excess burden associated with all real-world forms of taxation—rather than stemming from a bias against the public sector as such.

The well-known problem of the fungibility of funds remains just as difficult as ever. In the end, evaluators must despair of knowing where the money *really* went when loan funds are used to finance a project that would have been built anyway. Devarajan, Squire, and Suthiwart-Narueput and I respond to this problem by urging greater vigilance in all projects and programs. If all of them pass the cost-benefit test, fungibility entails no serious cost. Because the actual level of vigilance is so bad in nearly every country, there are many paths that would lead to its improvement. The paths urged by Devarajan, Squire, and Suthiwart-Narueput (public expenditures reviews) and by me (institutionalizing the widespread employment of cost-benefit analysis) are both sensible directions.

Observations

The four issues examined here go to the core of applied welfare economics. As one delves into them, one discovers conceptual subtleties that often go unrecognized. One also realizes how rudimentary are many of the practices that have become habitual among project evaluators.

It all goes back to the mantra of project evaluation courses—comparing the situation "with" and "without" the project. If only evaluators could do this really well, all or even most of the time! Part of the concern over whether a project should be undertaken by the public sector may stem from the ethos of our time—a time in which budgetary problems have been a prime concern, in which issues of efficiency in government have come to the fore, and in which privatizations and other forms of downsizing have helped make at least some governments significantly smaller than they were in the recent past. I believe that evaluators as well as donors can be sensitive to these considerations without appealing to any antigovernment bias simply by specifying the relevant alternatives much more carefully than they have in the past, by using in their discounting a properly calculated economic opportunity cost of capital, and by greatly improving the quality of the work they do, using tools of applied welfare economics that have been available all along. In short, the fundamental conceptual framework need not be modified, but there is plenty of room to improve what is actually done within that framework. Devarajan, Squire, and Suthiwart-

Narueput have performed a useful service by pointing out several important directions in which such improvement is needed.

Note

Arnold C. Harberger is professor of economics at the University of California, Los Angeles, and Gustavus F. and Ann M. Swift Distinguished Service Professor Emeritus at the University of Chicago.

1. For at least a couple of decades, it was common to speak of social project evaluation and of social opportunity costs of labor, foreign exchange, and capital. In the 1970s or early 1980s, economists began referring to economic project evaluation and to the economic opportunity costs of productive inputs and outputs. This change in terminology involved no change in meaning that I can detect. I believe it was motivated by a desire to avoid people's misinterpreting the concepts via an association of the term "social" with things like "social policy," "social services," "social problems," and so on. That linkage was never really present, and the use of the adjective "economic" better reflects this fact.

POINT-COUNTERPOINT

An occasional series of articles and comment, presenting differing views.

EDUCATION VOUCHERS IN PRINCIPLE AND PRACTICE: A SURVEY

Edwin G. West

An education voucher system exists when governments make payments to families that enable their children to enter public or private schools of their choice. The tax funded payments can be made directly to parents or indirectly to the selected schools; their purpose is to increase parental choice, to promote school competition, and to allow low-income families access to private schools. Some opponents predict that vouchers will destroy the public system, aggravate poverty, and foster segregation. Others fear that voucher-receiving independent schools will be regulated out of recognition.

The main purpose of this article is to examine the recent emergence of voucher systems as an interesting phenomenon in its own right. The evidence summarized relates to voucher systems operating in twenty countries, provinces, and states. The typical "funds-follow-the-child" voucher system, in which governments subsidize "schools of choice" in strict proportion to enrollment, appears to be the favorite form. This type of voucher has been adopted by developing countries—notably Bangladesh, Belize, Chile, Colombia, Guatemala, and Lesotho—as well as by industrial countries such as Poland, Sweden, the United Kingdom, and the United States. Much of the recorded experience with such programs is pertinent to the longstanding theoretical debates on the desirability of voucher systems.

A tax funded education voucher in the broadest sense is a payment made by the government to a school chosen by the parent of the child being educated; the voucher finances all or most of the tuition charged. The system introduces competition among public schools and between public and private schools; and it enables schools to offer diverse educational packages to meet the different preferences of parents.

The voucher systems discussed here apply to education up to and including high school and are funded through tax revenues (for a discussion of vouchers in higher education and privately funded voucher systems, see West 1996). First,

The World Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 83–103

however, it is important to understand the rationale for the basic interventi
that calls upon taxpayers to finance education.

The Rationale for State Intervention

In economics the three most quoted normative reasons for state interventi in education are to protect children against negligent parents, to internalize ben eficial "externalities," and to ensure equality of opportunity. Compulsory edu cation laws are generally regarded as satisfying the first argument for state in tervention. The externalities argument, to be completely persuasive, needs th support of evidence that externalities really exist and are positive at the mar gin—that is, that people outside the family unit are willing to pay for extra uni of education beyond what parents would purchase. In the absence of formal systematic evidence, most writers simply assume, explicitly or implicitly, th positive marginal external benefits do exist.

The third argument for intervention—the need to ensure equality of opportu nity—reflects concern about the distributional implications of purely private prov sion. Richer parents are likely to spend more than poorer parents to educate the children, just as they spend more on cars, homes, and clothes. The view that childre life chances should not depend on the wealth of their parents or the fortuitous ci cumstances of the community in which they live is widely accepted. The prospect upward mobility, of ensuring that one's children will be better off, has been a ke stone of political support for the public school system in the past.

This "equality" argument for intervention depends on the assumption th governments are best equipped to supply the appropriate institutions. But a publi system that confines children to schools nearest their home or within admini tratively determined attendance zones can actually reduce mobility. And whe the quality of public education is better in middle-class zones than elsewhe upward mobility is obviously blocked. In other words, the public system c often narrow a child's options, forcing the child to attend an inferior scho when a superior one may be physically within reach. One of the arguments f vouchers is that they enable families to break through these obstacles to gi equal opportunity a genuine chance.

The Rationale for Voucher Systems

The goal of all voucher plans—to provide families with maximum choi within a decentralized and competitive system of schools—embodies four prin ciples: consumer choice, personal advancement, the promotion of competitio and equal opportunity. Consumer choice, in education, equals parental choic parents choose schools for their children by virtue of their parental authorit and are thus, in a fundamental sense, the real consumers of education. Under

voucher plan, government serves the consumers of education—parents—rather than the suppliers of education—schools.

The second principle, that of personal advancement, is rooted in the conviction that people want to shape their own destinies. The opportunity to choose and to decide stimulates interest, participation, enthusiasm, and dedication. Many government programs—for example, Social Security, welfare, health programs, student loans—directly subsidize the individual recipients with funding for services among which they can select. Social security recipients, for example, can spend their checks however they choose. The goal of educational vouchers is to extend this principle to education.

The third principle, the stimulation of competition applies here because public schools are usually monopolies. The objective of vouchers is to challenge them to compete—with each other and with private schools—through reducing costs, increasing quality, and introducing dynamic innovation.

The fourth principle—the goal of equality of opportunity—underlying the rationale for vouchers is a logical outcome of the other three and is expressed in the objective of increasing access to private schools. This goal is embodied particularly in those "selective," or targeted, voucher schemes that give low-income families greater access to private schools, schemes that have been advocated by Oakland (1994) and Becker (1995). Oakland concludes that a case can be made for some redistribution in the provision of social services generally but suggests that redistribution is better accomplished by extending the welfare system to provide the poor with vouchers for selective government services such as education. This is in preference to the usual system whereby higher levels of government supply lower levels with grants that vary with the levels of local wealth and income. Although fiscal considerations are a factor in Becker's recommendation, he advocates a targeted system primarily "because the bottom quarter or so of the population are most in need of better education" (p. 11). He quotes studies that not only demonstrate the superior performance of private over public schools in the United States, but also show that "students from disadvantaged backgrounds tend to gain the most from attending private schools." This result, he observes, is not surprising "in light of the more extensive choices available to middle class and rich students" (p. 12).

Studies comparing the performance of public with private schools in developing countries generally appear to match those in the United States. Analysis, for instance, by Lockheed and Jimenez (1994) of private and public secondary schools in five developing countries revealed that private schools have a significant advantage both in student achievement and in unit costs.

Different Applications of the Voucher Principle

Under most tax-funded voucher systems, education is compulsory up to a legal school-leaving age, but parents are free to choose among alternative sup-

pliers of the compulsory service. Compared with an education tax rebate, vouchers help even those who pay little in direct taxation.

With vouchers children are not assigned to schools by attendance zones or any other criterion of the school system. Instead, vouchers enable parents to select a school for their children among any eligible and participating schools, public or private. In the most common application of the voucher principle, known as "funds follow the child," government funding is directed straight to the school chosen by the parent. Because it has no other direct government subsidy, each school is thus in competition with every other school for students. Good schools attract many students, redeem many vouchers, and prosper. Inferior schools, avoided by parents, are stimulated to improve or must close down.

In practice, tax-funded voucher systems operate under many different regulatory rubrics. They may include government inspection of schools receiving the vouchers. They may also operate only under the condition that the teachers are licensed by the government. Vouchers may be available to all families or to low income families. The value of the vouchers can also be made to vary inversely with income, so that poorer families receive vouchers worth more than those received by richer families. A variant of the funds-follow-the-child arrangement is a system of chits, given to each parent, cashable only by appropriately designated schools, who then return their vouchers to the relevant government authority and receive the cash value, which they use to pay expenses such as staff salaries. The value of the chit could be equal to, or somewhat less than, per student government expenditure in public schools. Finally, vouchers might provide access to private schools only, public (government) schools only, or to both public and private schools

Selective Vouchers

Selective vouchers can be restricted to families receiving less than a given income level. Such vouchers can of course be found outside the context of education. They have been used for housing, for health, and—perhaps the best example for these purposes—for food, in the United States federal government's food stamp program. The federal government uses an income test to determine eligibility for food stamps. Recipients use the stamps instead of cash to buy groceries. The grocery stores then return the stamps to the federal government and receive cash in return. This method is similar to the "chits" version of education vouchers described above. But whereas black market operations seriously threaten the food stamp system, the school voucher largely avoids this problem because it is quite difficult to transfer (sell) the rights to the education obtained.

Selective vouchers can be allocated on the basis of gender as well as income. In Bangladesh, for instance, vouchers are supplied exclusively to females in grades six through ten.

pen Enrollment and Charter Schools

It is sometimes contended that the objectives of vouchers can largely be hieved exclusively within the public sector. This argument involves the so-called pen enrollment system," wherein the family can choose public schools across tensive geographic areas. In practice, however, disproportionate applications enroll in a popular school lead administrators to declare it to be full. Unpopu-r schools, therefore, are not faced with serious costs of undercapacity and [illegible] continue to survive such weak competition.

Another potentially interesting scheme is the relatively new phenomenon of arter schools. These are decentralized and fairly autonomous institutions that erate under contract or charter to an authorized public body. If a charter hool does not attract and keep its students, it will go out of business and its arter will be revoked.

Because government subsidizes the charter school in direct proportion to its rollments, the voucher principle is at least partially respected because "funds llow the child"; for the principle to be fully respected, private schools would so have to be eligible to receive the grants. Nevertheless the charter school [illegible]des some alternative to the one public school in a child's administration ea to which he or she is usually assigned. In urban areas, moreover, parents y be able to choose between charter schools themselves. Further details of [illegible] institutions in the United States and Europe are provided by West (1996).

oucher Systems in Operation

Table 1 summarizes voucher systems for primary and secondary education have been implemented in twenty countries, states, or provinces around the e in the 1990s. Typically these voucher systems are the funds follow the kind, in which governments subsidize schools in strict proportion to en ments.

Space does not allow extended discussion of each entry in table 1. Five coun-s have therefore been selected for brief comment here, as case studies they cast some light on the arguments for and against vouchers reviewed in the section. The countries are Chile, Colombia, Puerto Rico, the United States waukee), and the United Kingdom.

hile

lowing the introduction of subsidized ("voucherized") private education Chile in 1980, the number of students attending private schools increased siderably. By 1988 private schools accommodated 30.4 percent of the el entary school population (compared with 14 percent in 1980) and 40.8 per nt of total secondary school registration (compared with 15.9 percent in 1980).

Table 1. *Education Vouchers: A Cross-Country Survey of Primary and Secondary Schooling*

Country	*Qualifying population*	*Coverage*	*Regulations and practices*	*Monetary value of voucher per student*
Bangladesh	Females grades 6–10	Selected localities	Public or private schools, minimum attendance and progress required	From $12 in grade 6 to $36.25 in grade 10
Belize	Elementary and secondary school attendees	75 percent of primary, 50 percent of secondary students	Strong government partnership with the churches	Not available
Canada: Province of British Columbia	Families patronizing independent schools	Denominational and secular private schools	Schools receiving vouchers have to have been established for 3 years minimum	30 percent of public school costs per student $500 in 1978
Canada: Province of Québec	Families patronizing independent schools	Mainly private secondary schools	Public inspection; teachers must have same qualifications as in public schools; same curriculum	60 percent of the costs of public schooling (80 percent for schools "in the public interest")
Canada: Province of Manitoba	Families patronizing independent schools	Private schools	Public inspection; teachers must have same qualifications as in public schools, same curriculum	Full-time equivalent capitation grants
Canada: Province of Saskatchewan	Families patronizing independent schools	Private schools	Curriculum, teacher qualifications, enrollment	55 percent of public school per capita cost
Canada: Province of Alberta	Families patronizing independent schools	Private schools	Curriculum, teacher qualifications, language requirements	50 percent of public school cost
Chile	Low-income elementary and secondary school attendees	More than one-third of total enrollments	Receiving schools can also charge fees	Average value in 1991: 4,359 pesos
Colombia	Low-income students [illegible]	Operational in 21[illegible]	Program participation renew-	$143 a year

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Japan	Children over 15 years old not covered by compulsory education laws	Private senior high schools	Private schools must submit financial statements to the foundation for the promotion of private schools	40 percent of the cost in private high schools covered by government
Lesotho	Elementary and secondary school attendees	Most schools	Government trains and appoints teachers, strong partnership with the churches	Not available
Netherlands	Children subject to compulsory education	[illegible]	State finance of schools for each religion where local demand demonstrated, secular private schools also state financed	Public and private schools are financed on a completely equal basis
New Zealand	All school-age children	All public schools, selected independent schools	Open enrollment system in a decentralized public sector; school autonomy strengthened via local parent elected boards	Teacher salary grants to independent schools amounting to 20 percent in 1993 with expressed intentions to raise it eventually to 50 percent
Poland	Families associated with one of the 56 sponsoring organizations, including the University of Warsaw	Private, [illegible] schools	Government approval required to open independent schools, a wide variety of curricula allowed in practice	Per capita subsidy level at 50 percent expenditure
Puerto Rico (until 1995)	Families with school-age children and incomes below $18,000	Public and private schools	Use of a lottery when demand for vouchers exceeds supply	$1,500
Sweden	Children subject to compulsory education	All municipal areas	Schools must follow national curriculum; supervision by the National Assembly of Education	At least 85 percent of per pupil cost in municipal schools

(Table continues on the following page.)

Table 1. *(continued)*

Country	*Qualifying population*	*Coverage*	*Regulations and practices*	*Monetary value of voucher per student*
United Kingdom	Low-income students with above-average ability	"Assisted Places" in private schools only	Participating schools must be approved by Education Department	$3,500 (approx.) a year on average (1992)
United States (1)	Low-income students in Milwaukee, Wisconsin; maximum 1,500 students	Private nonsectarian schools	Participating schools must limit voucher students to 65 percent of the student body	$2,900 a year (1994)
United States (2)	Low-income students in Cleveland City School District (K to 3rd grade), 11,864 students in the current pilot scheme	Public and private schools from K-3rd grade up through 8th grade; religious schools are included	Vouchers in the form of checks payable to the parents of "scholarship" students	Based on tuition charged by private schools of choice; normally around $3,000 a student

Source: Chile—Winkler and Rounds (1993); Martinez Sanchez (1993); Sweden, New Zealand, The Netherlands—OECD (1994); Japan—Lynn (1986); U.S. Milwaukee—McGroarty (1994); Puerto Rico—The Heritage Foundation (1995); United Kingdom—U.K Department of Education (1992); World Bank (1995); Flew (1995); Poland—Glenn (1995); Canada—Easton (1988); Bangladesh, Belize, Lesotho, Colombia, Guatemala—National Program of Self-Administration for Educational Development (PRONADE) (October 1996).

The Chilean reforms were described by the government as a move toward decentralization. Public schools were transferred to the municipalities, and a new subsidy law provided for the allocation of resources on a per-pupil basis and on equal conditions to both private and municipal schools. A "student performance examination" called Programa de Rendimiento Escolar (PER) operated between 1982 and 1984. This program lasted only two years because it encountered political difficulties. The Sistema de Medición de Calidad de Educación (SIMCE) national test followed in 1988. It indicated that the quality of education was significantly higher in the subsidized private educational establishments than in the municipal schools (with the exception of one group). The reforms were followed by an increase in the average number of years of schooling among the Chilean population, including the lower-income groups.

Economic recession has brought some setbacks in recent years, notably a reduction in the real value of the voucher, but to offset this partially, the new private schools have been allowed, since 1993, to charge fees for their services. This provision enables parents voluntarily to pay additional sums to their school with the object of trying to maintain or increase educational quality. Municipal primary (elementary) schools are not allowed to charge fees.

Colombia

A voucher system was introduced in 1992 and by 1994 was operating in 216 municipalities, serving 90,807 low-income students in 1,789 schools. The vouchers, worth on average about $143, were issued to students entering the sixth grade. An early examination of the program confirmed that, as intended, the vouchers were being successfully allocated exclusively to poor families.

The voucher system was introduced primarily to respond to the shortage of places in public secondary schools in Colombia, where 40 percent of the secondary schools are privately owned. The vouchers help poor students gain access to the private schools; simultaneously, the vouchers benefit the public secondary schools by reducing overcrowding.

The Colombian experience recalls that of Vermont in the United States, where approximately 95 percent of the state's 246 communities have no public secondary schools. The communities choose instead to pay tuition for their students to attend either private high schools or public high schools in another town. This program has been in place for more than a century "to enable small and geographically distant communities around the state to provide high school education for students without incurring the expense of building their own public schools" (Walberg and Bast 1993, p. 109).

Puerto Rico

Puerto Rico's governor, Pedro Rosello, signed a voucher plan into law in September 1993, which was limited to families earning below a given income.

The vouchers, worth $1,500, have been portable between public schools, as we
as from private to public and public to private schools; religious schools we
also included.

Preliminary evidence appears to refute opponents' predictions that a vouch
program would ruin the public school system. Of the 1,809 vouchers awarde
in the fall of 1993, 1,181 were used by students to transfer from one publ
school to another, 317 to move from private to public schools, and 311 to sh
from public to private schools.

Following opposition and litigation from the teachers' unions, who argue
that it was unconstitutional to spend vouchers at schools affiliated with rel
gions, the Supreme Court of Puerto Rico on November 30, 1994, ruled (5-
that the scholarship program allowing low-income students to attend the scho
of their choice violated Puerto Rico's constitution. The court, however, perm
ted the program to continue until the end of the school year (1995). Meanwhi
Governor Rosello and other supporters have promised to try to find a way
continue the program.

Milwaukee

One of the most striking examples of a successful voucher system for th
poor is found in Milwaukee, Wisconsin, in the United States. Pioneered larg
by Democrat representative "Polly" Williams in 1990, the plan originally pe
mitted up to 1,000 low-income students to use state funds ($2,967 for the 1994
95 school year; the amount is adjusted annually) to attend a private, nonsecta
ian school of their choice.

The Milwaukee program began operation in 1990 with 300 children using
vouchers at six private schools. Five years later (1995) 832 students attende
one of eleven participating private schools. The Milwaukee plan has been op
posed by various educational establishment groups, including the State Scho
Board Association and the Wisconsin Congress of Parents and Teachers, Inc
This opposition has probably influenced the administrative restrictions that hav
accumulated recently. Thus in 1994 the state legislature set a ceiling on th
program of 1.5 percent of Milwaukee's 100,000 school-aged population, or 1,500
students. The private schools participating in the program must limit voucher
students to 49 percent of their student body, which limits the number of places
available. Since the program's inception, the lack of space has meant that more
students have been turned away than have been accepted into the program. In
consequence, spaces are apportioned by lottery (McGroaty 1994).

The Milwaukee scheme, though small, warrants attention because it is the
only source of hard evidence on the effects of vouchers in the United States.
Comments on the program's performance have been based on the annual re
ports of Professor John F. Witte, the state-selected outside evaluator. His first
reports led some critics to complain that the participating schools suffered ex
cessive attrition (dropouts) and that achievement tests were biased because the

mothers of the families using vouchers had a higher average high-school completion rate than mothers of students who did not use vouchers. These complaints were later rebutted by McGroaty (1994).

Those findings of Witte's evaluations that are unambiguously positive, meanwhile, combat three of the popular fears or predictions about the voucher program, discussed in greater detail in the next section. The first is the suspicion that vouchers will help individuals who are not poor and who therefore need help least. Witte's evidence shows, on the contrary, that "choice families" are among the poorest of the poor. Their average income in 1994 was $11,625—half the income level of the average family with children in Milwaukee's public schools.

The second commonly expressed fear is that vouchers will lead to segregated and antisocial schools. Evidence supplied by Witte shows instead that the Milwaukee program fosters diversity and that no participating school has been teaching cultural supremacy or separation. "The student bodies of participating [voucher] schools vary from schools that are almost all one minority race, to racially integrated schools, to schools that have used the Choice program to diversify their almost all white student bodies" (Witte and others 1995, p. 15).

The third fear—that voucher schools will skim off the "cream" of the student crop—is countered by Witte's finding that "the program is offering opportunities for a private school alternative to poor families whose children were not succeeding in school. This is a positive outcome of the program" (Witte and others 1995, p. 16).

Other positive conclusions from Witte's reports include the finding of high parental involvement, once in the system, and high parental satisfaction with the program—in particular, that it increased learning and discipline. "Respondents almost unanimously agreed the program should continue" (Witte and others 1995, p. 17).

The case for vouchers rests also on the argued need to weaken the public school monopoly or, in other words, to promote competition. But when competition is introduced, those suppliers who initially lose, or expect to lose, customers will, in self-defense, act to lift the quality of their services. Applied to our education context, five years of the Milwaukee plan is more than enough time for the threatened public schools to have improved under the pressure of new voucher competition. And insofar as vouchers can take some credit for inducing the improvements in tested achievement that have in fact occurred over the years 1990–96 (in public and private schools), findings of no current difference in achievement growth between public and voucher (choice) schools do not unambiguously imply that vouchers have failed to improve efficiency.

The future of vouchers in the United States obviously will be influenced not only by official annual reports, but also by the assessments and responses of the parents. The fact that demand for voucher places in Milwaukee currently well exceeds supply could already be pressuring politicians to allow more families to participate.

United Kingdom

In 1981 the Assisted Places Scheme was established in the United Kingdom with the aim of providing a ladder of opportunity for able but poor students. Under the scheme today, low-income parents can obtain assistance with tuition fees for an independent school if the school has been approved by the Department of Education and Science.

By 1995 about 29,800 students were using these selective vouchers at 29. specified independent schools in England (there is a separate system for Scotland). About 5,000 new pupils enter the program every year, mostly at the age of eleven or thirteen.

The English experience raises two questions that have implications for the general debate on vouchers discussed in the next section. First, why—in view of the government's stated wish to encourage competition and "market discipline"—is the Assisted Places method so limited in coverage (see U.K. Department of Education 1992)? Second, why are the places limited mainly to able pupils who exhibit the potential for high academic achievement, when such pupils can expect a higher than average lifetime income whether or not they are in Assisted Places? The contrast with Milwaukee's selective voucher, where the low income students designated for help have not been succeeding in school, is striking.

The voucher principle has also been extended in the United Kingdom to further education and (prospectively) to nursery schooling. Further education colleges (similar to community colleges in the United States) have recently been reestablished as autonomous institutions independent of their former local governments. A new system of "learning agreements"—effectively, individual contracts between a college and a student, specifying the precise qualifications aimed for—enables government funding to follow the student to the college of his or her choice. At the preschool level, the Department of Education declared in 1995 that it was about to extend free entitlement for all four-year-olds to good quality private, as well as public, nursery education (World Bank 1995, p. 4 [illegible]). Currently the initiative has been limited to two pilot schemes in East Anglia. The plan, however, is to be extended to all four-year-olds in April 1997 (The Economist 1996).

The Current Debate on Vouchers

As the case for parental choice and competition has gained in popularity, the criticism of those antipathetic to vouchers has increased in intensity. Debate has focused on the potential effects of vouchers on the public benefits connected with education; the possibilities for damage to the quality of public schools on the one hand or to the identity and autonomy of private schools on the other; their impact, if any, on poverty; the issue of windfall gains for the middle class; and the possible effect of a voucher system on the government's administrative costs.

Vouchers and the Market Place

Some view vouchers primarily in terms of a free market that vouchers would encourage. They then see this as a prime example of "economic man" sacrificing social welfare to his selfish pursuit of individual material gain. But economists have long abandoned narrow assumptions about self-interest. As Becker (1993, p. 385–6) observes: 'Behavior is driven by a much richer set of values and preferences. [My] analysis assumes that individuals maximize welfare *as they conceive it*, whether they be selfish, altruistic, loyal, spiteful, or masochistic.' A pertinent example is the objections made by the members of the Polish Civic Educational Association in the late 1980s to the national school system inherited from the collapsed Communist regime. Their position was that they wanted to maximize welfare as *they* as individuals saw it, as a welcome change from having welfare defined and imposed by totalitarian authorities or highly centralized bureaucracies. The type of institutions they demanded were nonstate (including religious) private or independent schools (Glenn 1995, p. 127).

A related argument by opponents of vouchers is that a free market would lead to discrimination on grounds of race or disability. Krashinsky (1986, p. 143) argues that vouchers could lead to racial segregation. The usual reply here is to quote Coleman's (1990) findings that segregation is in fact greater in public than in private schools. Shanker and Rosenberg (1992) suggest in the same vein as Krashinsky that profit making schools would reject difficult-to-educate children under a voucher system. Lieberman (1991a) found, on the contrary, that the single largest U.S. group of for profit schools serves the disabled. Blum (1985), meanwhile, provides evidence that urban private schools maintain a higher level of discipline than do public schools.

Another common argument against vouchers is that parents cannot be expected to make sound choices for their children (Bridge 1978, Carnegie Foundation 1992, Levin 1991, Wells and Crain 1992). Others reply that parents simply need some initial experience (hitherto denied them) with making such choices in order to become more adept. A second response is that, in a democracy, any serious impediments to decisionmaking by parents will show up also at the ballot box when they choose political representatives to make decisions on education. A third response is to quote empirical studies demonstrating rational choice of their children by parents who themselves have only modest amounts of education (Fossey 1994).

A further concern—that vouchers (or tax credits) for education might introduce fraudulent practices—is put forward by Murnane (1983), who draws an analogy with food stamps in the United States. Experience there, he observes, shows that unscrupulous parties make claims for fictitious individuals. Schmidt (1995) shows that serious shortcomings of fraud and dishonesty are already present in the public school system. Moreover, and to reiterate, the school voucher largely avoids the black market problem because it is difficult to transfer the rights to education.

Public versus Private Benefits from Education

It is generally accepted that a child's education provides not only private benefits to the family (mainly by prospectively increasing income), but also public benefits (positive externalities). The latter include poverty reduction, economic growth, and the pursuit of common values (see Krashinsky 1986). The economic model supporting this argument is that of "joint supply." One classic example of it is wool production: wool cannot be produced without simultaneously producing meat, and vice versa. Furthermore, a switch from one breed of sheep to another is likely to improve the wool production at the expense of meat, or the converse. Similarly, so the argument goes, the cost of more or improved public benefits from education shows up in fewer, or worse quality private benefits, introducing an interesting trade-off problem. The public benefits are quite distinct from the private. Thus the inducement to an orderly society that educated citizens bring is one example of a public benefit. The increase in expected lifetime income that education bestows on students is, in contrast, an example of a purely private benefit.

Some economists object to free choice of schooling through a voucher system because they believe families will not trade off private for public benefits but will allocate their expenditures on the basis of their private benefits exclusively. In other words, the valuation that others in society place upon the education of one's child will be neglected, and public benefits will suffer relative to private benefits—the well-known "public good problem."

Proponents of that view, such as Krashinsky (1986) and Levin (1991), claim that public schools have a unique ability to produce the "common values" just mentioned. But this claim also is now contested. Cohn (1979), for instance, observes that, in practice, public schools in the United States have successfully resisted attempts to homogenize their procedures, so that "a student in one school district might receive an entirely different set of common values than his counterpart in another school district" (p. 262). Nevertheless the belief that public schools possess an absolute advantage in producing the "public good" benefits remains strongly entrenched among educationists as well as among some economists.

Private schools are direct producers of externalities (Hettich 1969), and they also generate them *indirectly* (West 1991). It is generally agreed that private schools are more efficient at producing private benefits, through more effective teaching of the basics, such as literacy. This is so partly because public schools are monopolies, while private schools have greater output per dollar because they experience competition. But literacy is also a public benefit, a necessary condition for communicating common values and fostering economic growth. This indirect assistance by private schooling to the production of such public benefits is at least as important as the direct production.

Krashinsky (1986) focuses on what he calls transaction costs, such as the costs of communication in obtaining the public benefits of education. His position is that these costs are too high for the government to contract out to private

Vouchers and Poverty Reduction

Krashinsky's implicit assumption is that the public school system benefits the poor in a way that is superior to any alternative. But low-income families are segregated residentially, and their children are typically allocated to the schools nearest their homes. If they want to choose a better public school in a middle-class area, they must purchase a home there, and the housing prices are usually beyond their means. Middle-class families, by contrast, can move more easily because they are less restricted financially. The result is that the public provision of schooling becomes heterogeneous, with the poor, on average, receiving the worst quality. Vouchers would help remove the barriers to mobility.

Friedman and Friedman (1980) insist that they too favor reducing poverty and promoting equal opportunity but argue that in both respects the voucher system would unmistakably improve things. They contend that liberty, equality of opportunity, and the reduction of poverty are complementary and not competitive goals of the voucher system. Their main argument is that lower-income families, trapped in large city ghetto schools, would benefit most from vouchers. "Are the supermarkets available to different economic groups anything like so divergent in quality as the schools?" they ask. "Vouchers would improve the quality of the public schooling available to the rich hardly at all; to the middle class, moderately; to the low-income class, enormously" (p. 169).

Windfall Gains for the Middle Class

Some opponents of vouchers focus on what they call the inequitable windfall gains for families (usually well-to-do) that customarily purchase private education. In other words costs to governments would increase if vouchers (or tax credits) are extended to rich private school clients not now financed by government (Gemello and Osman 1983). Seldon (1986) points out, however, that *net* costs to government could fall depending on the value of the voucher as a proportion of per capita public school costs. The government savings would occur according to Friedman (in Seldon 1986, p. 20), if the voucher value was [illegible] percent of public school costs. The reasoning is that the economies effected by migrants from public to private schools, who would now cost the government 25 percent less than before, would offset the cost of the windfall gain to accustomed users of private schools. (Clearly, because a strong argument put forward by voucher supporters is that private schools can deliver at lower costs than public, their case looks more consistent if they demand vouchers at values less than 100 percent of average per pupil costs in public schools.)

The windfall gains problem could also be handled by making vouchers subject to tax. But selective voucher systems, restricted to low-income families would be even more effective—indeed, such selective vouchers would automatically prevent high-income families now patronizing private schools from enjoying the windfall gains.

Regulatory Threats to Private School Identity

A potential drawback to vouchers has recently been suggested by strong believers in the philosophy of freedom who want to see more competition in schooling but fear that voucher systems would seriously threaten the autonomy of independent schools. Currently the most articulate and influential spokesman for this point of view in the United States is Sheldon Richman (1994). In his words: "It is likely that before schools could accept vouchers, they would be required to meet a raft of standards that before long would make the private schools virtually indistinguishable from public schools" (p. 83). Voucher initiatives that insisted on zero regulation would stand no chance of acceptance, Richman says, because, "as the opposition would inevitably point out, the voucher plan would appear to authorize appropriation of 'public' money to institutions not accountable to 'public authorities'" (p. 84). In the same vein, Gary North (1993) argues: "We will have federal guidelines operating in every voucher-using school, equal opportunity policies and quota systems of every kind, teacher hiring and firing policies, racially and religiously mixed student bodies. There will be a whole army of federal bureaucrats, not to mention state bureaucrats policing every 'private' school" (p. 149).

Friedman has always separated three levels of issues: first, whether schooling should be compulsory; second, whether it should be financed privately or by the government, and third, how it should be organized. His position has been that whatever one's views may be on the first two issues, a voucher scheme would produce a better and a more effective organization than the present one—that is, vouchers remain a superior alternative to a system of schools run and financed by government. Like North and Richman, Friedman sees benefits also in eventually removing compulsion and government finance, but he is primarily concerned with the question of how to get there from here. Vouchers, he believes, are still a practical transitional measure (Friedman 1993).

As for the threat of a government regulatory take over of private schools, Anderson (1993) points out that these institutions do not have to accept vouchers with all their strings. Others argue also that the recipients of vouchers can and will lobby their government against heavy regulation. Lieberman (1991a, p. 6), meanwhile, argues that the more likely cause of increased regulation will be the practical objections to funding both public and private schools while closely regulating only the former. Consequently, Lieberman observes, supporters of vouchers must argue that to approach parity, what is needed is the reduction of the regulation of public schools, not an increase in the regulation of private schools.

The Costs of Implementation

A common concern about the administrative costs of implementing a voucher system is whether the size of the bureaucracy necessary to oversee the total system will have to expand significantly. Wilkinson (1994) finds no reason to

believe that costs such as those associated with monitoring student attendanc
and quality of education should be any higher for private than for public schools
school quality can be overseen by periodic inspections in the same way as it is i
public schools. Even in the unlikely event that administrative costs did rise, such
an increase would be more than offset by the savings realized given the evidence
cited above that private schooling generally costs less than public. Tax-funde
vouchers in the countries described in table 1 are typically valued at consider
ably less than the public school per capita cost; the Milwaukee plan, for in
stance, supplies students with vouchers worth about half of the public scho
cost. It is highly improbable that additional administrative costs could equ
such a huge differential. Indeed, a strong argument for governments to use vouch
ers is the need in these days of budgetary stringency to economize on publ
spending.

Final Comments

The main purpose of this article has been to provide information on the theo
and practice of education vouchers throughout the world and to summariz
briefly the principal points raised in current academic and political debates
the issue. Absence of real world evidence has until recently hampered disc
sion—indeed, until recently, has been adduced by several writers to de
strate that vouchers were not desirable. But emerging evidence (see table
suggests otherwise, and this may well be due to changing circumstances.

During the last two decades governments have become increasingly unwil
or unable to continue to raise the share of public expenditure spent on edu
tion. The prime focus has switched accordingly to attempts to obtain hig
output from given expenditure levels. The use of vouchers valued at much l
than 100 percent of the cost per pupil in public schools has already been succ
ful in Sweden, Milwaukee (United States), and Poland, and may become a po
lar way of economizing. Economists, meanwhile, see the key role in such eff
ciency gains to be the gradual removal of the current monopoly struct
education.

Some have argued (for example, Carnoy in this issue) that a central diff
with the voucher plan is that, even if it is limited to the poor, it "diverts at
tion from an even greater problem: the much larger investment that socie
need to make in low-income children if they hope to overcome the effects
poverty on learning." Carnoy does not indicate, however, how this large
vestment is to be financed. The voucher system could help him, because s
stantial resources could be made available for his program by switching
means of vouchers) to the significantly lower cost of education supply that pri
vate schools offer.

It is too early to reach firm general conclusions about the effectiveness of
vouchers. There are only twenty entries in the table, and these show a wi

Holmes, Mark. 1988. "The Funding of Independent Schools." In James E. Lam and C. Benjamin, eds. 1988. *Public Policy and Private Education in Japan.* London: Macmillan

Krashinsky, Michael. 1986. "Why Educational Vouchers May Be Bad Economics." *Teachers College Record* 88(2, Winter):139–51.

Levin, Henry M. 1991. "The Economics of Educational Choice." *Economics of Education Review* 10(2):137–58.

Lieberman, Myron. 1991a. "The Case for Voluntary Funding of Education." Policy Study 37. The Heartland Institute, Chicago, Illinois. Processed.

Lieberman, Myron. 1991b. *Privatization and Education Choice.* New York: St. Martin Press.

Lockheed, Marlaine E., and Emmanuel Jimenez. 1994. "Public and Private Secondary Schools in Developing Countries." HRO Working Paper 43. World Bank, Washington, D.C. Processed.

Lynn, Robert. 1986. "The Japanese Example of Indirect Vouchers." *Economic Aff...* 6(6):25–27.

Matte, Patricia, and Antonio Sancho. 1993. "Primary and Secondary Education." In Chr... tian Larroulet, ed., *Private Solutions to Public Problems.* Santiago, Chile: Inst... Libertad Y Desarrolo

McGroaty, Daniel. 1994. "School Choice Slandered." *The Public Interest* 117(Fall):...

Murnane, Richard J. 1983. "The Uncertain Consequences of Tuition Tax Credits: An Analy... of Student Achievement and Economic Incentives." In Thomas James and Henry M. Levin, eds., *Public Dollars for Private Schools.* Philadelphia, Pa.: Temple University Pre...

National Program of Self-Administration for Educational Development (PROSADE), October 1996.

North, Gary. 1993. "Education Vouchers: The Double Tax." *Freeman* 43(6):72–8...

OECD (Organization for Economic Cooperation and Development). 1994. *School: A Ma... of Choice.* Paris.

Oakland, William. 1994. "Fiscal Equity." In John E. Anderson, ed., *Fiscal Equaliz... for State and Local Government Finance.* London: Praeger Publishers.

Richman, Sheldon. 1994. *Separating School and State.* Fairfax, Virginia: The Futur... Freedom Foundation.

Schmidt, Paul. 1995. "Looking the Other Way." *Education Week* 14(21):23–2...

Seldon, Arthur. 1986. *The Riddle of the Voucher.* Hobart Paperback 21. London: Inst... of Economic Affairs.

Shanker, Albert, and Brian Rosenberg. 1992. "Politics, Markets, and American Scho... Rejoinder." In P. R. Kane, ed., *Independent Schools, Independent Thinkers.* San Fr... cisco, Calif.: Jossey-Bass.

Tucker, Allyson M., and William F. Lauber, eds. 1995. *School Choice Programs: Wh... Happening in the States.* Washington, D.C.: Heritage Foundation.

U. K. Department of Education. 1992. *Choice and Diversity.* London.

Walberg, Herbert J., and Joseph L. Bast. 1993. "School Choice: The Essential Refo..." *CATO Journal* 13(1).

Wells, A. S., and R. L. Crain. 1992. "Do Parents Choose School Quality or School St... A Sociological Theory of Free Market Education." In P. W. Cookson, ed., *The C... Controversy.* Newbury Park, N.Y.: Corwin Press.

West, Edwin G. 1991. "Public Schools and Excess Burdens." *Economics of Educ... Review* 10(2, July).

———. 1996. "Education Vouchers in Practice & Principle. A World Survey" (full... port). Available on request from the author at Carleton University, Department of E... nomics, 1125 Colonel By Drive, Ottawa K1S 5B6, Canada. Processed.

Wilkinson, Bruce. 1994. *Educational Choice: Necessary But Not Sufficient.* Montreal, Canada: Institute for Research on Public Policy.

Winkler, Donald R., and Taryn Rounds. 1993. "Municipal and Private Sector Response to Decentralization and School Choice." HRO Working Paper 8. World Bank, Washington, D.C. Processed.

Witte, John F. and others. 1995. *Fourth Year Report: Milwaukee Parental Choice Program.* Madison, Wisc.: Department of Public Instruction.

World Bank. 1995. "Education Reform in the United Kingdom: Autonomy with Accountability." Human Development Department. Washington, D.C. Processed.

IS PRIVATIZATION THROUGH EDUCATION VOUCHERS REALLY THE ANSWER? A COMMENT ON WEST

Martin Carnoy

Government-financed school vouchers promise to improve consumer choice while still providing the public funding needed for families to invest adequately in their children's education. But politically and in practice, the choice that vouchers offer can mean many things, from eliminating neighborhood residency as a condition for attending a public school to allowing groups of parents and teachers to form their own public schools (charter schools in the United States, for example). And it can mean an educational system that is in part publicly funded but that is privately owned and operated with no public regulation. Choice through vouchers can also produce undesirable side effects that more than offset their optimistic promises.

In this context, E. G. West presents a particular version of the case for vouchers. Although he allows for a range of possibilities, he puts primary emphasis on the *privatization* of education. This version of choice has to be distinguished from one that limits increased choice to public schools and from the use of vouchers to improve education for the poor. The three emphases are different, mainly in what they contend are the social benefits and costs of various degrees of choice. Whereas the case presented by West may appeal at an abstract level, I argue that the evidence supporting it is at best mixed and possibly negative.

The Foundation of West's Claims

West's argument for vouchers rests on a foundation of three claims that bear careful scrutiny. The first is that privately managed education is inherently more effective and more cost effective than is public education. The second is that a privatized education system is more efficient than public education in improving the social mobility of youngsters from low-income families. The third is that the social costs of privatizing a public educational system are minimal.

If private education is not both more effective and more cost-effective than public education, extending vouchers to unregulated private schools might in-

The World Bank Research Observer, vol. 12, no. 1 (February 1997), pp. 105–16

crease the competition in false claims or symbols of success, but it would not support West's claim of "reducing costs, increasing quality, and introducing dynamic innovation" (West, p. 85). Further, if the meaning of effectiveness is redefined to include the production of outputs other than measured learning and not in the public "charter," private schools could be more effective than public schools and could increase some families' private welfare but might not necessarily enhance the public good. For example, in the United States and many other countries, private schools can offer religious education, which is not legal in public schools. Religious training for their children is considered a benefit by some families. Yet the use of the tax funds for religious education could be contrary to the political principles of the vast majority of the national community and could increase the level of conflict over community goals.

And unless private education *is* more effective, its relevance to the principles of choice and personal advancement put forth by West is also less clear. West (p. 87) argues that a publicly monopolized education system restricts "good schooling—that, in practice, even under an "open enrollment system" of public school choice, popular, more effective public schools declare themselves "full and poor schools continue to operate without threat of further competition. Thus extending vouchers to private schools would increase choice by expanding the number of *good* school places available. This implies that private schools are both more effective and more cost-effective than public schools, and that with private schools free to attract students, either they would not declare themselves full or, seeing a chance to make money, additional *effective* private schools would enter the market and expand the number of places available. The argument assumes the existence of (potentially) large numbers of private schools that are more effective than public schools. (Otherwise, new private schools would not emerge or would be unable to compete in the longer, or even medium, run with public education.)

It is possible that the inclusion of private schools in the voucher plan would expand the choice options, even with equally effective public and private schools. That possibility occurs, as noted above, if private schools can offer outputs that public schools cannot, such as religious education or guaranteed admission to university or a job in a particular company.

West contends that a privatized system of education has even greater positive consequences for poor families than for wealthier families, because it would equalize the distribution of educational quality and educational outcomes among families with different incomes. He also argues that extending vouchers to private schooling would have few and only minimal negative consequences. In West's discussion, these two claims are related, because potential downsides to privatization are closely linked to negative distributional effects, namely, harm to the public school system and windfall gains to middle- and upper income groups. West contends that existing voucher plans have resulted in little, if any, negative impact on public education and perhaps have even had a positive result because they have increased competition. Further, he suggests that negative dis-

tributive effects could be handled by limiting vouchers to low-income families, as is the case in some countries. But this caveat is inconsistent with his more general claims for privatization. Limiting vouchers to low-income families restricts choice, which contradicts the choice principle for vouchers. In any case, the main reason to argue for a voucher plan limited to low-income families is to protect against the likelihood that a free market in education for all families would produce greater educational inequality than a public school system.

If the main issue is improving the education of low-income children, privatization through vouchers has both positive and negative aspects. On the positive side, it draws attention to the poor quality of many public schools in low-income areas and may benefit more highly motivated families. Gary Becker (1995) argues that "the best voucher system is limited to poor families" because "the bottom quarter or so of the population are most in need of better education, and the poor are most likely to benefit from competition by private schools." I agree that if a voucher system were limited to the poor, the poor could benefit, although not necessarily because access to private education would raise their test scores. Studies comparing Catholic with public education in the United States suggest that Catholic schools give students from low income families a better self concept, but not much higher test scores (Bryk, Lee, and Holland 1993). The objective of such a voucher system, however, would be to alleviate poverty, not to increase school choice or privatize schooling. Politically, this distinction is important. On the negative side, such a voucher plan diverts attention from the much larger investment that societies need to make in low income children to overcome the effects of poverty on school achievement.

Is Private Education More Effective?

As West has noted, a number of voucher plans are in operation, but only two—Chile and Milwaukee, Wisconsin—have data that attempt to compare academic results for voucher pupils attending public and private schools.

Evidence from Chile

The Chilean plan, implemented by the military regime in 1980 as part of an overall free-market package, provided for fully subsidized deregulated private schools, which competed head-on for pupils with deregulated public schools in all metropolitan neighborhoods. But because few rural areas have any private schools, children in these areas (more than a fourth of Chile's municipalities) attend only public schools (Winkler and Rounds 1993). The Chilean plan revoked teachers' contracts and eliminated the teachers' union as a bargaining unit, so by 1983 public schools could hire and fire teachers without regard to tenure or union contracts. At the same time, all schools were released from the rigid structure defining the national curriculum and national standards.

return to public schools (the attrition rate was 30 percent). In addition, several of the original private schools closed, forcing voucher students to return to public schools. As a result only the most motivated students in the successful private schools remained after four years.

Another recent study, by Greene, Peterson, and Du (1996), estimates significantly higher test scores for third- and fourth-year voucher students compared with students who had applied to the voucher program but had not been accepted in the lottery; this was true, however, only in the case that socioeconomic background differences are not corrected for. They claim that their study makes a true correction for selection bias because it compares voucher students with an identical group of students who had qualified for the voucher program and thus were equally motivated and met the low socioeconomic requirements of the voucher plan. Witte (1996) argues, however, that the Greene-Peterson-Du methodology is never fully specified and appears to be flawed and that once socioeconomic status is controlled for, the differences in test scores are no longer significant.

On average, the private elementary schools in Milwaukee's choice program seem to have provided instruction at lower cost than did the public schools, at least in the first years of the program. The cost data are not strictly comparable, however, because private school tuition fees do not reflect the real cost per pupil. But as teachers in the private schools were paid about a third less than those in public schools and class size was approximately the same, the implication is that the unit costs for private schools are about a third less than those for public schools. Because test scores do not differ across comparable groups, this means that private schools were more effective than public schools. However, salaries of teachers in private schools have apparently risen more rapidly than those of public-school teachers. At today's $4,375 voucher, per-pupil costs in private schools appear to be approximately the same as those in Milwaukee's public elementary schools.

Is Private Education Better for Low-Income Students?

In a review of the European situation, John Ambler (1994, p. 470) argues that in Britain, France, and the Netherlands "the primary negative effect of school choice is its natural tendency to increase the educational gap between the privileged and the underprivileged." The Netherlands is especially interesting because about 70 percent of pupils attend fully subsidized private schools. There are waiting lists to get into the "better," more selective schools, which charge fees on top of the voucher subsidy. To make lower-income, "disadvantaged" pupils more attractive, the government provides them with a voucher that is 25 percent more than the average. Even so, the students who end up in municipal (public) schools are mainly those lower-income pupils (with the larger vouchers) who still cannot attend the more desirable schools.

e private schooling is widely available to lower-income families in *ur-*
cipalities, but, as argued above, analysis of test scores suggests that
m such families may have gained least from the privatization process.
West claims that privatization helped expand Chile's secondary edu-
condary education did expand rapidly in the 1980s), and therefore
social mobility for low-income Chileans, there is no evidence that
education would not have expanded even more rapidly through pub-
on, as it did in many countries of Latin America and Asia in the 1970s
s. After Haiti, Brazil has the most privatized secondary education sys-
tin America, and access by low-income pupils is highly limited.
bia's voucher plan is also targeted at students from low-income fami-
/es those who could not qualify academically for free public secondary
, but whose families are willing to pay the tuition at private schools
above the voucher. There appears to be little "substitution" effect,
ans that the plan really does expand the number of places available to
ne students, and in its first years, the plan is allowing most of these
academically excluded students to attend traditional private paro-
ols that have excess capacity and relatively high academic standards.
dents undoubtedly benefit from the voucher plan. But once the capac-
sting schools is reached, private schools of much lower quality may
ousiness. This is already a problem, and the Colombian government
ed in to regulate these schools (Calderon 1996). Is expanded low-
ivate education paid for by government funds superior in any way to
(and possibly low-quality) public education costing about the same?
could have a high payoff for the "best" group of excluded low-income
p to the point at which they can be absorbed by existing higher quality
hools, but beyond that, payoffs could decline rapidly.

a Downside to Privatization?

rgues specifically that there is little to lose from privatizing education
to gain. Although there is no convincing evidence that the gains exist,
not mean privatization should not be tried. That is, unless it could also
net social loss.
ssible loss is greater inequality. Evidence suggests that privatizing edu-
ough vouchers for all pupils may produce a more unequal distribution
onal quality and choice than under a predominantly public system. If
to attend a private school is not equally distributed across the popu-
is the case in Chile and in Europe), private (and even public-school
ons) might widen the divide between urban and rural and between
d higher-income students (Levin 1995). Chile's voucher plan appears
idened the gap between high- and low-income students in terms of test
hout increasing the overall level of academic achievement. It also re-

return to public schools (the attrition rate was 30 percent). In addition, several of the original private schools closed, forcing voucher students to return to public schools. As a result only the most motivated students in the successful private schools remained after four years.

Another recent study, by Greene, Peterson, and Du (1996), estimates significantly higher test scores for third- and fourth-year voucher students compared with students who had applied to the voucher program but had not been accepted in the lottery; this was true, however, only in the case that socioeconomic background differences are not corrected for. They claim that their study makes a true correction for selection bias because it compares voucher students with an identical group of students who had qualified for the voucher program and thus were equally motivated and met the low socioeconomic requirements of the voucher plan. Witte (1996) argues, however, that the Greene-Peterson-Du methodology is never fully specified and appears to be flawed and that once socioeconomic status is controlled for, the differences in test scores are no longer significant.

On average, the private elementary schools in Milwaukee's choice program seem to have provided instruction at lower cost than did the public schools, at least in the first years of the program. The cost data are not strictly comparable, however, because private school tuition fees do not reflect the real cost per pupil. But as teachers in the private schools were paid about a third less than those in public schools and class size was approximately the same, the implication is that the unit costs for private schools are about a third less than those for public schools. Because test scores do not differ across comparable groups, this means that private schools were more effective than public schools. However, salaries of teachers in private schools have apparently risen more rapidly than those of public-school teachers. At today's $4,375 voucher, per-pupil costs in private schools appear to be approximately the same as those in Milwaukee's public elementary schools.

Is Private Education Better for Low-Income Students?

In a review of the European situation, John Ambler (1994, p. 470) argues that in Britain, France, and the Netherlands "the primary negative effect of school choice is its natural tendency to increase the educational gap between the privileged and the underprivileged." The Netherlands is especially interesting because about 70 percent of pupils attend fully subsidized private schools. There are waiting lists to get into the "better," more selective schools, which charge fees on top of the voucher subsidy. To make lower-income, "disadvantaged" pupils more attractive, the government provides them with a voucher that is 25 percent more than the average. Even so, the students who end up in municipal (public) schools are mainly those lower-income pupils (with the larger vouchers) who still cannot attend the more desirable schools.

In Chile private schooling is widely available to lower-income families in *urban* municipalities, but, as argued above, analysis of test scores suggests that pupils from such families may have gained least from the privatization process. Although West claims that privatization helped expand Chile's secondary education (secondary education did expand rapidly in the 1980s), and therefore increased social mobility for low-income Chileans, there is no evidence that secondary education would not have expanded even more rapidly through public education, as it did in many countries of Latin America and Asia in the 1970s and 1980s. After Haiti, Brazil has the most privatized secondary education system in Latin America, and access by low-income pupils is highly limited.

Colombia's voucher plan is also targeted at students from low-income families. It serves those who could not qualify academically for free public secondary schooling, but whose families are willing to pay the tuition at private schools over and above the voucher. There appears to be little "substitution" effect, which means that the plan really does expand the number of places available to low-income students, and in its first years, the plan is allowing most of these otherwise academically excluded students to attend traditional private parochial schools that have excess capacity and relatively high academic standards. Those students undoubtedly benefit from the voucher plan. But once the capacity of existing schools is reached, private schools of much lower quality may open for business. This is already a problem, and the Colombian government has stepped in to regulate these schools (Calderon 1996). Is expanded low-quality private education paid for by government funds superior in any way to expanded (and possibly low-quality) public education costing about the same? Vouchers could have a high payoff for the "best" group of excluded low-income students up to the point at which they can be absorbed by existing higher quality private schools, but beyond that, payoffs could decline rapidly.

Is There a Downside to Privatization?

West argues specifically that there is little to lose from privatizing education and much to gain. Although there is no convincing evidence that the gains exist, that does not mean privatization should not be tried. That is, unless it could also lead to a net social loss.

One possible loss is greater inequality. Evidence suggests that privatizing education through vouchers for all pupils may produce a more unequal distribution of educational quality and choice than under a predominantly public system. If the choice to attend a private school is not equally distributed across the population (as is the case in Chile and in Europe), private (and even public-school choice plans) might widen the divide between urban and rural and between lower- and higher-income students (Levin 1995). Chile's voucher plan appears to have widened the gap between high- and low-income students in terms of test scores without increasing the overall level of academic achievement. It also re-

duced government efforts to improve schooling, in part because the free market was expected to increase achievement. Only after the central government intervened with increased in-service training, new curriculum standards, and technical assistance to low-income schools did achievement levels appear to increase across the board.

Better private schools for middle-class pupils may produce higher test scores than do public schools, but mainly by keeping out hard-to-manage or less-motivated pupils. As Winkler and Rounds (1993) report, parents in Chile select schools mainly on the basis of the characteristics of the pupils that already attend those schools. Willms and Echols (1992) found the same behavior in Scotland, where a choice plan allowed parents to select from among public schools. But private schools are not uniformly exclusive or effective; some accept (and keep) less-motivated or less able youths who score poorly. Why would parents send their children to such schools? One answer to this question is that because some private schools are associated with a higher social stratum and higher test scores (not necessarily because they are more effective, but because they limit enrollment), their reputation could rub off on all private schools. Winkler and Rounds (1993) cite an unpublished experiment in which changing the school name from Spanish to English and introducing school uniforms in public schools (techniques used by private subsidized schools in recruiting pupils) increased enrollments in public schools.

West could claim that well-informed parents would soon withdraw their children from less effective private schools and reenroll them in public schools. But that might not be the case if vouchers had a negative effect on public schools. If, for example, private schools attract the better students from public schools, the latter will increasingly be seen as suitable only for low-income or slow students. West argues that there is no such effect, but he falls back on evidence from voucher plans that cater only to less able, low-income youths, such as those in Milwaukee and Colombia. West cites Witte's assertion that private schools in Milwaukee did not draw the "cream of the crop" or help those who needed help the least as evidence that private schools would not do either under a generalized voucher plan. But the voucher plans in Milwaukee and Colombia are restricted to those very low-income pupils who need help the most. Other students could not qualify for vouchers. Witte was simply reporting that the objectives of this restricted plan were being met and that those low-income students who were dissatisfied with the public schools and so participated in and stayed with the program were, by and large, satisfied. But recent attempts by the State of Wisconsin to offer vouchers to the school population as a whole have prompted concern in Milwaukee's African-American community, which backed the targeted voucher plan, that the broader eligibility will favor higher-income families (Miner 1995).

School choice, after all, is also affected by the costs of transportation and by how much parents know about various schools. West implicitly assumes that neither is a factor in determining the alternatives available. Yet studies of choice

plans in both education and health care suggest that the location of alternatives is crucial and that the higher the income of the consumer, the more quality, rather than location, plays the key role (Elmore 1990). Those who are better off can provide or pay for private transportation for their children to a wider range of school locations, so the same voucher for all means more restricted access for the poor.

Similarly, less-educated families are less likely to search out and use information on the quality of educational alternatives. This does not mean that less-educated families are not aware of the quality of their children's education, particularly if that education is especially poor or especially good. Rather it suggests that choice systems "require substantial information on alternatives in order to support effective decision-making . . ." and that "education is a complex service in which even those with more sophistication must rely heavily on word of mouth and interpretation of incomplete data on school quality" (Levin 1995, p. 10). Proponents of privatization sometimes characterize the view that less educated parents have less information as "elitist," but all the empirical evidence suggests that it reflects reality. "Studies of parental information on choice show that less educated, minority, and lower-income parents are often unaware of their choice options and are not cognizant of differences among schools of choice . . . [and] studies have consistently shown that . . . the lower the socioeconomic status of a family, the less likely that it will 'choose' the school attended by its offspring . . ." (Levin 1995, pp. 11-12).

Thus introducing competition between private and public schools (and even among public schools) may actually exacerbate inequalities in the educational system. By increasing the incentive to rely on public relations techniques (such as changing to an English name) that have nothing to do with the school's effectiveness, the free market preys on a less educated clientele. The idea that increased competition in the production of complex, difficult-to-define products, such as education, automatically increases social welfare is simply wrong. This is especially true where private profit is the driving force behind the competition. That is why governments have been called upon to regulate private education and protect the public interest in the voucher plans cited here.

It is also possible that privatizing education could reduce the incentives of the politically powerful middle class to push for more resources to education. Although these parents would allegedly still have an interest in raising the average amount of the voucher to offset the cost of private (or public) education, it is also possible that private schools could raise their fees as the vouchers increased in value. When vouchers were introduced in Sweden, for example, the private schools that previously charged tuition raised their fees by 9 percent (Miron 1993).

Furthermore, a privatized system could affect the distribution of public money, shifting funds from teachers' compensation to private school operators. Between 1983 and 1990, teachers' salaries in Chile dropped sharply, especially in the more privatized secondary school system. The decline was partly the result of a

steadily lower voucher value, but it was also related to the profits marked for the owners of the schools. Although many would applaud lower salaries for teachers as one way to make education less costly, this approach is not necessarily more cost-effective. Teachers are crucial to improving education. If labor markets work properly, lower salaries for teachers will attract less talented people. Because teaching conditions are usually more difficult in schools with low-income pupils, the effect of reducing teacher pay could have its greatest impact on such schools.

In an analysis of Great Britain's 1980 public-school choice law, Willms and Echols (1992, p. 347) concluded that the moderate gains that may have been achieved through contextual effects when children went to better schools may have had high costs for the educational system as a whole:

> Schools serving pupils in disadvantaged areas will be receiving incorrect signals; many of the them will lose pupils to higher SES [socioeconomic status] schools despite effective teaching practices. Some high SES schools will also receive incorrect signals, because many parents are choosing these schools even though their performance is mediocre or poor when compared with schools with similar social class intakes.

Privatization or Public Improvement?

There is no persuasive evidence that private schools are more effective than public schools and the evidence that they are more cost-effective is mixed. Moreover, not every income group's welfare is raised through increased choice. Even if satisfaction and educational attainment were raised in absolute terms, a loss in relative position could make lower-income groups feel worse off than before. Unlike income, the absolute level of education or even quality of education attained is probably less important than the relative amount and quality of education an individual or group attains, since the payoff associated with a given level of education declines rapidly as a large fraction of the population attains it (Carnoy 1972).

West would limit vouchers to low-income families and restrict privatization to the poor as a means of equalizing an otherwise unequal system. This is a good answer to the distribution of choice problem, but it does not address the question of effectiveness. Private schools do not seem to improve the achievement of pupils from low-income families any more than they do for the pupil population as a whole. Even so, data from Milwaukee and Colombia suggest that a voucher program limited to the poor could make low-income families feel better off than they would if the present system of education were left untouched. Yet limiting vouchers to poor families contradicts the broader argument of privatization advocates.

The voucher question boils down to how societies feel politically about the responsibilities and role of the state in education. The political judgment that

the public sector is not capable of adequately providing education but should finance it supports the view that vouchers for private education are appropriate even if they are unlikely to improve the average level of learning. Many of the arguments against privatizing education can be applied to public choice plans, such as Scotland's. So implementing a voucher plan that allows parents to send their children to public schools outside of their neighborhoods or to regulated "charter" schools that serve niche markets within public school districts is also a political decision, one that is bound to favor certain groups more than others. If a society wants to offset the effects of family poverty on children's education, it could use vouchers limited to the poor as one element of an antipoverty program. If a society is convinced that both dissatisfied and more inert parents are better served by efforts to improve neighborhood public schools than by efforts to provide exit options to the dissatisfied, it should focus on improving public education.

I would like to believe, with Professor West, in a panacea that could make everyone learn more without investing enormous time and effort in improving children's nutrition, home lives, and the way *all* schools deliver knowledge. Our task as educators and social reformers would be that much simpler. Unfortunately, vouchers tend to divert attention from the overall complexity of the learning problem rather than providing a real solution.

Note

Martin Carnoy is professor of education and economics at Stanford University

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

[illegible], John. 1994. "Who Benefits from Educational Choice? Some Evidence from Europe." *Journal of Policy Analysis and Management* 13(3):454–76.

[illegible], Gary. 1995. "Human Capital and Poverty Alleviation." HRO Working Paper 52. World Bank, Washington, D.C. Processed.

[illegible], Anthony, Valerie Lee, and Peter Holland. 1993. *Catholic Schools and the Common Good*. Cambridge, Mass.: Harvard University Press.

[illegible] Z., Alberto. 1996. "Voucher Programs for Secondary Schools: The Colombian Experience." Human Capital Development Working Paper 66. World Bank, Washington, D.C. Processed.

[illegible], Martin. 1972. "The Political Economy of Education." In Thomas LaBelle, ed., *Education and Development in Latin America and the Caribbean*. Los Angeles: UCLA Latin American Center.

[illegible], Martin, and Carlos Torres. 1994. "Educational Change and Structural Adjustment: A Case Study of Costa Rica." In Joel Samoff, ed., *Coping With Crisis: Austerity, Adjustment, and Human Resources*. Paris: ILO/UNESCO.

Comite tecnico asesor del dialogo nacional sobre la modernizacion del la educacion chilena 1994. *Los Desafios del la Educacion Chilena frente al Siglo XXI.* Santiago de Chile: Editorial Universitaria.

Elmore, Richard. 1990. "Choice as an Instrument of Public Policy: Evidence from Education and Health Care." In William Clune and John F. Witte, eds., *Choice and Control in American Education.* Vol. I. Philadelphia, Pa.: Falmer Press.

Espinola, Viola. 1993. "The Educational Reform of the Military Regime in Chile: The System's Response to Competition, Choice, and Market Relations." Ph.D. dissertation, University of Wales, United Kingdom. Processed.

Greene, Jay P., Paul Peterson, and Jiangtao Du. 1996. "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation." University of Houston and Harvard University, John F. Kennedy School of Government, Cambridge, Mass. Processed.

Levin, Henry. 1995. "Educational Choice and Educational Equity: Perspectives on the Chilean Case." School of Education, Stanford University, Stanford, Calif. Processed.

Miner, Barbara. 1995. "Rifts Grow in Voucher Movement." *Rethinking Schools* [illegible] (1, Fall):3, 26.

Miron, Gary. 1993. *Choice and the Use of Market Forces in Schooling.* Stockholm: University of Stockholm, International Institute of Education.

Prawda, Juan. 1993. "Educational Decentralization in Latin America: Lessons Learned." *International Journal of Educational Development* 13(3):253–64.

Rounds Parry, Taryn. 1994. "The Impact of Decentralization and Competition on the Quality of Education: An Assessment of Educational Reforms in Chile." University of Georgia, Department of Political Science, Athens, Ga. Processed.

Tsang, Mun. 1996. "Private and Public Costs of Schooling in Developing Nations." In Martin Carnoy, ed., *International Encyclopedia of Economics of Education.* [illegible] Oxford, U.K.: Pergamon Press.

Tsang, Mun, and W. Taoklam. 1992. "Comparing the Cost of Government and Private Primary Education in Thailand." *International Journal of Education and Development* 12(3):177–90.

Willms, J. Douglas, and Frank Echols. 1992. "Alert and Inert Clients: The Scottish Experience of Parental Choice of Schools." *Economics of Education Review* 11(4, December):339–50.

Winkler, Donald, and Taryn Rounds. 1993. "Municipal and Private Sector Response to Decentralization and School Choice: The Case of Chile, 1981–1990." World Bank, Latin America and Caribbean Region, Technical Department, Washington, D.C. Processed.

Witte, John F. 1996. "Reply to Greene, Peterson, and Du: 'The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation.'" University of Wisconsin-Madison, Institute of Public Affairs. Processed.

Witte, John F., Chris Thorn, and Kimberly Prichard. 1996. *Fifth Year Report: Milwaukee Parental Choice Program.* Madison, Wis.: State Department of Public Instruction.

Globalization and Inequality, Past and Present

Jeffrey G. Williamson

The late nineteenth and late twentieth centuries shared more than globalization and economic convergence. The trend toward globalization in both centuries was accompanied by changes in the distribution of income as inequality rose in rich countries and fell in poor ones. Between one-third and one-half of the rise in inequality since the 1970s in the United States and other member countries of the Organization of Economic Cooperation and Development (OECD) has been attributed to global economic forces, about the same as a century earlier. It appears that the inequality produced by global economic forces before World War I was responsible in part for the retreat from globalization after the war. What does this retreat imply for the future? Will the world economy once again retreat from globalization as the rich OECD countries come under political pressure to cushion the side effects of rising inequality?

Economic growth after 1850 in the countries that now belong to the Organization for Economic Cooperation and Development (OECD) can be divided into three periods: the late nineteenth century belle epoque, the dark middle years between 1914 and 1950, and the late twentieth century renaissance. The first and last epochs were characterized by rapid growth: economic convergence as poor countries caught up with rich ones; and globalization, marked by trade booms, mass migrations, and huge capital flows. The years from 1914 to 1950 are associated with slow growth, a retreat from globalization, and economic divergence. Thus history offers an unambiguous positive correlation between globalization and convergence. When the pre–World War I years are examined in detail, the correlation turns out to be causal: globalization was *the* critical factor promoting economic convergence (Williamson 1996a).

Because contemporary economists are now debating the impact of the forces of globalization on wage inequality in the OECD countries, the newly liberalized Latin American regimes, and the East Asian "tigers," it is time to

ask whether the same distributional forces were at work during the late nineteenth century. A body of literature almost a century old argues that immigration hurt American labor and accounted for much of the rise in income inequality from the 1890s to World War I. The decision by a labor-sympathetic Congress to enact immigration quotas shows how important the issue was to the electorate. An even older literature argues that cheap grain exported from the New World eroded land rents in Europe so sharply that landowner-dominated continental parliaments raised tariffs to protect domestic growers from the impact of globalization. But nowhere in this historical literature had anyone constructed data to test three contentious hypotheses with important policy implications:

Hypothesis 1: Inequality rose in resource-rich, labor-scarce countries such as Argentina, Australia, Canada, and the United States. Inequality fell in resource-poor, labor-abundant agrarian economies such as Ireland, Italy, Portugal, Scandinavia, and Spain. Inequality was more stable among the European industrial leaders, including Britain, France, Germany, and the Lowland countries, all of whom fell in between the rich New World and poor Old World.

Hypothesis 2: If the first hypothesis is true, a second follows: these inequality patterns can be explained largely by globalization

Hypothesis 3: If this second hypothesis holds, then these globalization induced inequality trends help explain the retreat from globalization between 1913 and 1950.

This article reviews the historical debate about the first globalization boom in the late nineteenth century and attempts to tie it to the current debate about the globalization boom in the late twentieth century. The two debates are strikingly similar. They also share a shortcoming in the empirical analysis: nobody has yet explored this issue with late nineteenth century panel data across poor and rich countries, and, with the important exception of Wood (1994), few have done so for the late twentieth century debate either (Burtless 1995, p. 813). Indeed, until very recently, most economists had focused solely on the American experience. The central contribution of this paper is to explore a database for the late nineteenth century that includes both rich and poor countries or, in the modern vernacular, North and South.

It appears that globalization did contribute to the implosion, deglobalization, and autarkic policies that dominated between 1913 and 1950. Indeed, during these years of trade suppression and binding migration quotas, the connection between globalization and inequality completely disappeared. It took the globalization renaissance of the early 1970s to renew this old debate.

Globalization and Inequality in the Late Twentieth Century

From 1973 through the 1980s, real wages of unskilled workers in the United States fell as a result of declining productivity growth and an increasing disparity in wages paid to workers with different skills (Kosters 1994; Freeman 1996). This difference was manifested primarily by higher wages for workers with advanced schooling and age-related skills. The same trends were apparent elsewhere in the OECD in the 1980s, but the increase in wage gaps was typically far smaller. The widening of wage inequalities coincided with the forces of globalization, both in the form of rising trade and increased immigration, the latter characterized by a decline in the skill levels of migrants (Borjas 1994). Trade as a share of gross national product in the United States increased from 12 percent in 1970 to 25 percent in 1990 (Lawrence and Slaughter 1993), while exports from low-income countries rose from 8 percent of total output in 1965 to 18 percent in 1990 (Richardson 1995, p. 34). These developments coincided with a shift in spending patterns that resulted in large trade deficits in the United States.

The standard Heckscher-Ohlin trade model makes unambiguous predictions: every country exports those products that use abundant and cheap factors of production. Thus a trade boom induced by a drop in tariffs or in transport costs will cause exports and the demand for the cheap factor to boom as well. Globalization in poor countries should favor unskilled labor; globalization in rich countries should favor skilled labor. Lawrence and Slaughter (1993) explored this wage inequality and concluded that there was little evidence to support the standard trade model explanation. Instead, the authors concluded that technological change was an important source of rising wage inequality. Hot debate ensued, with no resolution in sight.

This strand of the debate stressed the evolution of labor demand by skill, ignoring the potential influence of supply. Borjas (1994) and Borjas, Freeman, and Katz (1992) took a different approach, emphasizing instead how trade and immigration augmented the supply of labor in the United States. They first estimated the implicit labor supply embodied in trade flows, since imported goods increase the effective labor supply in the importing country. Similarly, exports imply a decrease in the effective labor supply in the exporting country. In this way, the huge U.S. trade deficit of the 1980s implied a 1.5 percent increase in the labor force, and because most of the imported goods used unskilled labor, it also implied a work force characterized by an increasing ratio of unskilled to skilled labor. Between the 1960s and the 1980s, an increasing proportion of immigrants to the United States were from developing nations, which meant that a far higher fraction were relatively unskilled just when there were more immigrants.

These shifts in the supply of labor produce the desired qualitative result for the purposes of this study—wage inequality between skill types. The quantitative result, at least in George Borjas' (1994) hands, also seems to be large: he estimates that 15 to 25 percent of the relative decline in the wages of high-school graduates compared with those of college graduates is attributable to globalization forces, of which trade accounts for one-third, immigration, two-thirds. He also estimates that 30 to 50 percent of the decline in the wages of high-school dropouts relative to the wages of all other workers is attributable to these same forces. Hatton and Williamson (1995; 1997) show that a century earlier, immigration was a far more dominant influence on U.S. inequality than was trade, and furthermore, that trade and migration influenced relative wages in practically every country involved in the globalization experience.

Thus far the discussion has focused mainly on the United States, perhaps because rising inequality and immigration have been greatest there. But the question is not simply why the demand for unskilled labor in the United States and even Europe was depressed in the 1980s and 1990s (Freeman 1995, p. 19), but whether the same factors were *stimulating* the relative demand for low-skill labor in developing countries. This is where Adrian Wood (1991, 1994, ch. 6; 1995b) enters the debate. Wood was one of the first economists to systematically examine inequality trends across industrial and developing countries.

Wood distinguishes three skill types: uneducated workers, those with a basic education, and the highly educated. The poor South has an abundance of uneducated labor, but the supply of workers with basic skills is growing rapidly. The rich North, of course, is well endowed with highly educated workers; its supply of labor with basic skills is growing slowly. Wood assumes that capital is fairly mobile and that technology is freely available. As trade barriers fall and the South improves its skills through the expansion of basic education, it produces more goods that require only basic skills, while the North produces more high-skill goods. It follows that the ratio of the unskilled to the skilled wage should rise in the South and fall in the North. The tendency toward the relative convergence of factor prices raises the relative wage of workers with a basic education in the South and lowers it in the North, producing rising inequality in the North and falling inequality in the South.

Wood concludes that the decline in the relative wages of less-skilled northern workers is caused by the elimination of trade barriers and the increasing abundance of southern workers with a basic education. He also dismisses skill-using technological change as a potential explanation for rising inequality because labor and total factor productivity growth both slowed during the period. Wood also argues that the pattern of increasing wage inequality in the North favors a trade explanation because there is no cross-country association between inequality trends and technological progress.

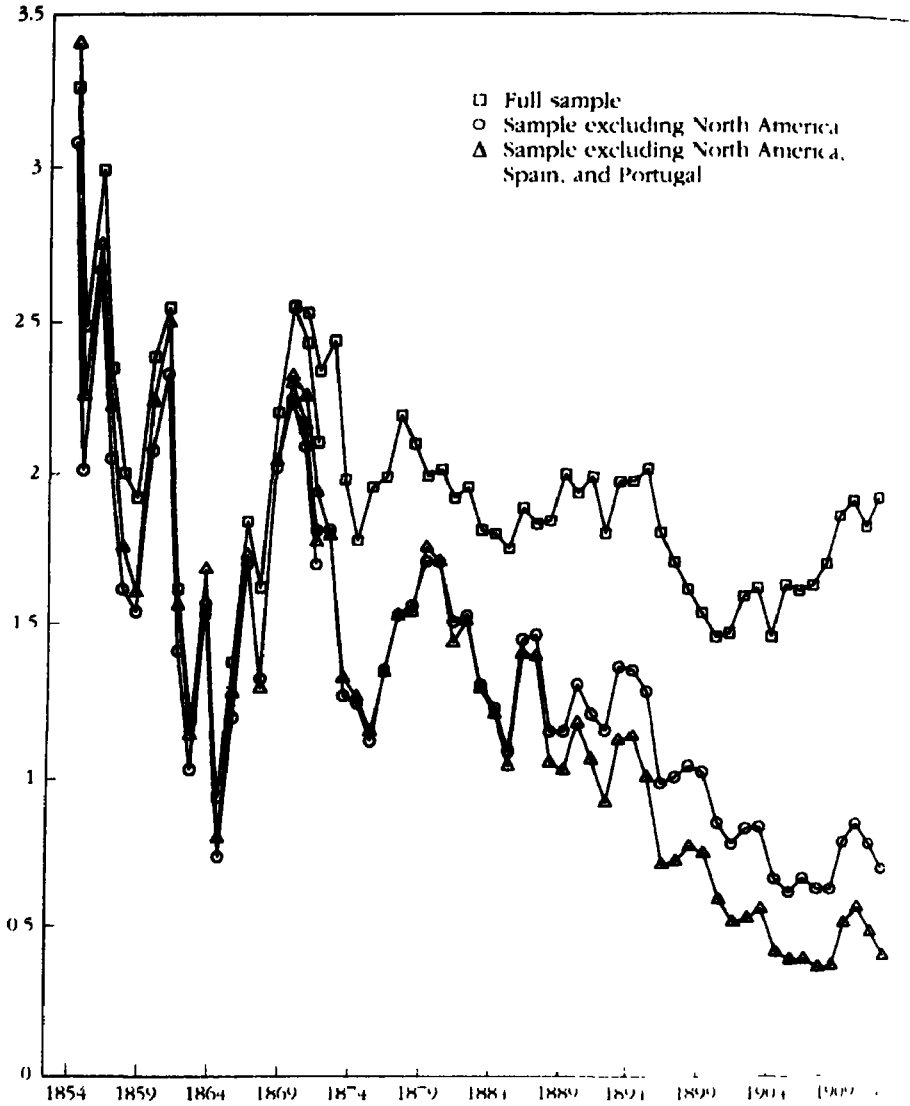
Wood's research has met with stiff critical resistance.[1] Since his book appeared in 1994, more has been learned about the link between inequality and globalization in developing countries. Economic theory argues that poor countries should become more egalitarian in the face of globalization, unless demographic or industrial revolution forces offset it. A recent review by Davis (1996) reports the contrary, and a study of seven countries in Latin America and East Asia shows that wage inequality typically did not fall after trade liberalization but rather *rose* (Robbins 1996). This apparent anomaly has been strengthened by other studies, some of which have been rediscovered since Wood's book appeared (Michaely, Papageorgiu, and Choksi 1991). Almost twenty years ago Krueger (1978) studied ten developing countries covering the period through 1972, and her findings were not favorable to the simple predictions of standard trade theory. Her conclusions have been supported by Bourguignon and Morrisson (1991) and by recent work on the impact of Mexican liberalization on wage inequality (Feenstra and Hanson 1995; Feliciano 1996). None of these studies is very attentive to the simultaneous role of emigration from these countries, however, leaving the debate far from resolved.

Globalization and Inequality in the Late Nineteenth Century

The spread between real wages from 1854 to 1913 in fifteen countries is shown in figure 1. The downward trend confirms what new-growth theorists call convergence, that is, a narrowing in the economic distance between rich and poor countries. The convergence is more dramatic when America and Canada—which were richer—or when Portugal and Spain—who failed to play the globalization game—are excluded. Convergence of gross domestic product (GDP) per worker hour has been reported elsewhere, based on Maddison's (1991) data. Most of this convergence was the combined result of the trade boom and the prequota mass migrations (Hatton and Williamson 1995; O'Rourke and Williamson 1994, 1995, 1996, and forthcoming; Taylor and Williamson 1997; Williamson 1995, 1996a).

Trade Issues

The late nineteenth century was a period of dramatic integration of commodity markets. Railways and steamships lowered transport costs, and Europe moved toward free trade in the wake of the 1860 Cobden-Chevalier treaty. These developments implied large trade-induced price shocks that affected every European participant. The drop in grain prices was the canonical case: wheat prices

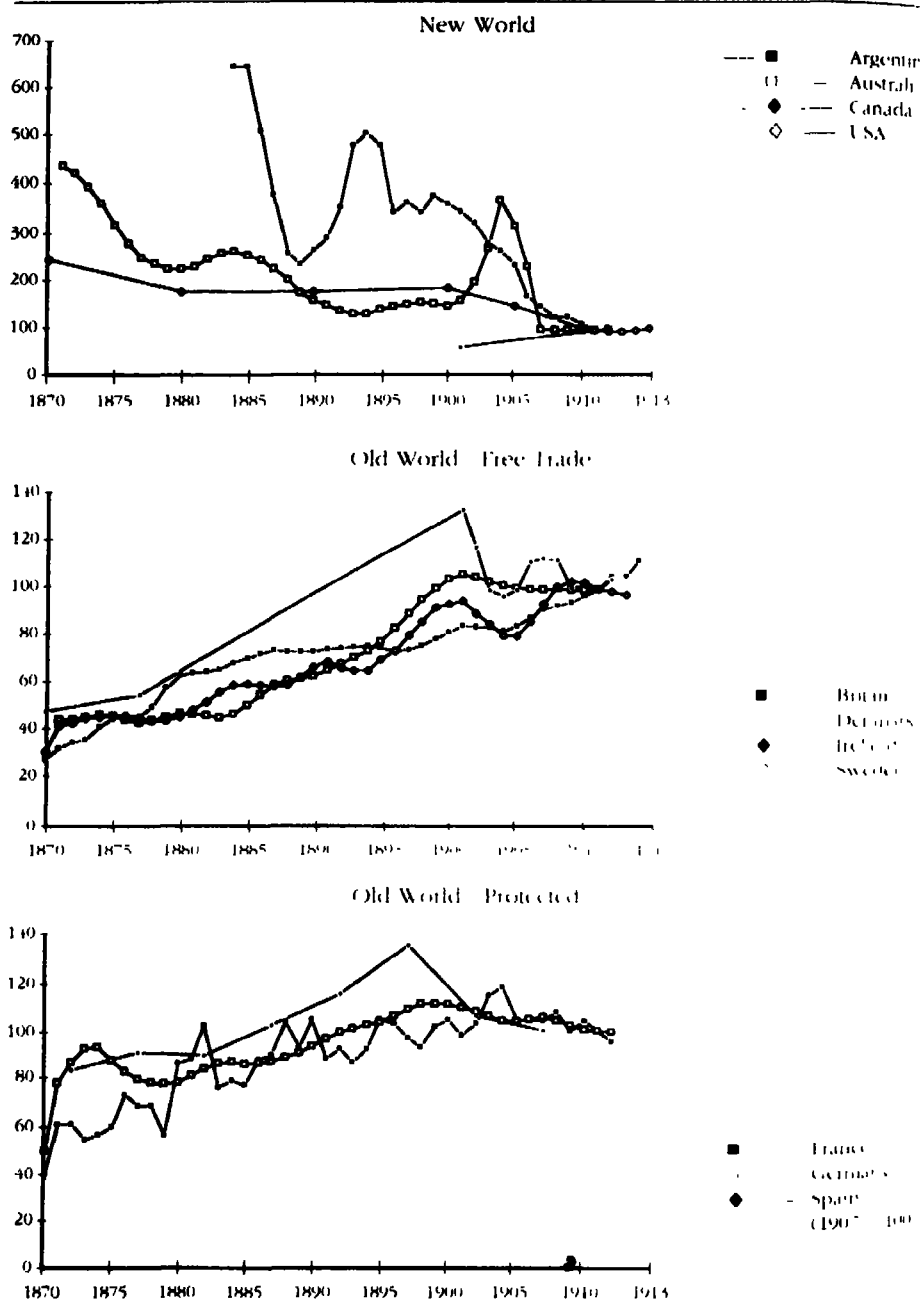
Figure 1. *Real Wage Dispersion, 1854–1913*

Note: Wage data are urban, male purchasing power parity adjusted.
Source: Williamson (1996a, figure 1).

in Liverpool were 60 percent higher than those in Chicago in 1870, for example, but they were less than 15 percent higher in 1912, a decline of forty-five percentage points. The commodity price differential declined by even more when the spread is measured from wheat-growing regions outside of Chicago. Furthermore, prices of all tradables, not just grain, were affected. It should be stressed that these globalization price shocks were *far* larger than those embedded in the infamous 1930 Smoot-Hawley tariff or any other U.S. tariff in the past century.[3] They were also larger than the decline in OECD tariff barriers induced by the General Agreement on Tariffs and Trade after the 1940s, events which triggered the globalization boom of the last quarter century. World Bank studies report that tariffs on manufactures imported by industrial countries fell from 40 percent in the late 1940s to 7 percent in the late 1970s, a drop of thirty-three percentage points. Wood (1994, p. 173) uses this example to advertise just how revolutionary world commodity market integration has been in recent decades, but even this spectacular drop is smaller than the forty-five percentage-point decline in trade barriers between 1870 and 1913 caused by improvements in transport.

The standard trade model argues that, as countries everywhere expand the production and export of goods that use their abundant (and cheap) factors relatively intensively, the resultant market integration would lead to an international convergence of factor prices. Under this theory, then, the late nineteenth century trade boom accounted for 10 to 20 percent of the convergence in GDP per worker hour and in the real wage.[4] It also had distributional implications for poor countries: it meant rising wages for unskilled workers relative to land rents and skilled wages. For rich countries, it meant that unskilled wages fell relative to land rents and skilled wages.

Migration Issues

The correlation between real wages or GDP per worker hour and migration rates is positive and highly significant. The poorest Old World countries tended to have the highest emigration rates, while the richest New World countries tended to have the highest immigration rates. The correlation is not perfect since potential emigrants from poor countries often found the cost of the move too high, and some New World countries restricted inflows of such migrants. But the correlation is still very strong. Furthermore, the effect on the labor force was very important, augmenting the New World labor force by almost 37 percent and reducing the Old World labor force by 18 percent (at least among the emigrant countries around the European periphery), much larger than U.S. experience in the 1980s. One estimate suggests that mass migrations explain about 70 percent of the real wage convergence in the late nineteenth century

Figure 2. *Ratio of Unskilled Wages to Land Values, 1870–1913*
(1911 = 100)

Source: O'Rourke, Taylor, and Williamson (1996, figs. 1, 2, 3).

(Williamson 1996a; Taylor and Williamson forthcoming). This estimate, in contrast with the contemporary debate about immigration in the 1980s, which focuses only on immigration into Europe and the United States, includes the total impact on rich receiving countries *and* poor sending countries.

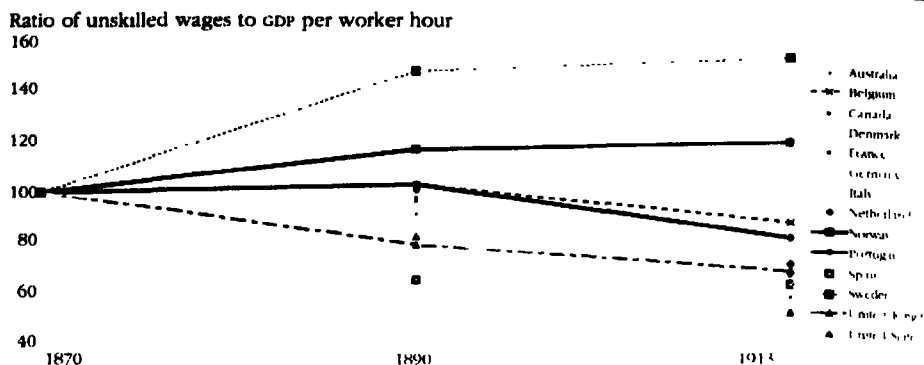
Because the migrants tended to be unskilled, and increasingly so toward the end of the century, they flooded the receiving countries' labor markets at the bottom of the skill ladder. Thus immigration must have lowered unskilled wages relative to those of skilled artisans and educated white-collar workers and relative to land rents. These immigration-induced trends implied increased inequality in rich countries, while emigration-induced trends must have moved in the opposite direction and reduced inequality in poor countries.

So much for plausible assertions. What were the facts?

Establishing the Facts, 1870–1913

How did the typical unskilled worker near the bottom of the distribution do relative to the typical landowner or capitalist near the top, or even relative to the skilled blue-collar worker and educated white-collar employee near the middle? The debate over inequality in the late twentieth century has fixed on wage inequality, but a century earlier, land and landed interests were far more important sources of income, so they need to be added to the inquiry. (I believe this is true throughout the developing world, certainly its poorer parts.) In any case, two kinds of evidence are available to document nineteenth century inequality trends so defined: the ratio of unskilled wages to farm rents per acre, and the ratio of the unskilled wage to GDP per worker hour. Everyone knows that farm land was abundant and cheap in the New World, while scarce and expensive in the Old World. And labor was scarce and expensive in the New World, while abundant and cheap in the Old World. Thus, the ratio of wage rates to farm rents was high in the New World and low in the Old. What everyone *really* wants to know, however, is how the gap evolved over time. Are the trends consistent with the predictions of the globalization and inequality literature? Was there, in Wood's language, relative factor price convergence in the late-nineteenth century, implying rising inequality in rich countries and declining inequality in poor countries? Figure 2 supplies some affirmative answers.

In the New World the ratio of wage rates to farm rents plunged. By 1913 it had fallen in Australia to a quarter of its 1870 level; in Argentina to a fifth of its mid-1880 level; and in the United States to less than half of its 1870 level. In the Old World the reverse occurred, especially where free trade policies were pursued. In Great Britain the ratio in 1910 had increased by a factor of 2.7 over its 1870 level, while the Irish ratio had increased even more, by a factor of 5.5. The Swedish and Danish ratios had both increased by a factor

Figure 3. *Normalized Inequality Levels, 1870–1913*

Note Inequality levels are normalized by setting the ratio of unskilled wages to GDP per worker hour at 1870 = 100

Source Williamson (1996b)

of 2.3. The surge was less pronounced in protectionist countries, increasing by a factor of 1.8 in France, 1.4 in Germany, and not at all in Spain.

Because landowners tended to be near the top of the income distribution pyramid,[7] this evidence confirms Hypothesis 1: inequality rose in the rich, labor scarce New World and fell in the poor, labor-abundant Old World. There is also some evidence that globalization mattered: countries that were open to trade absorbed the biggest distributional changes; those that retreated behind tariff walls sustained the smallest distributional changes.

So much for wage-rental ratios. What about the ratio of the unskilled worker's wage (w) to the returns on *all* factors per laborer, or GDP per worker hour (y)? Changes in w/y measure changes in the economic distance between the working poor near the bottom of the distribution and the average citizen in the middle of the distribution.

Figure 3 summarizes the wide variance across the fourteen countries in the sample. Powerful Danish and Swedish equality trends establish the upper bound 1913 = 244 (the index rises above 100); powerful Australian and U.S. *in*equality trends establish the lower bound, 1913 = 53 (the index falls below 100).[8]

An alternative way to standardize these distributional trends is simply to compute the annual percentage change in the index between 1870 and 1913, which ranges from +0.97 and +0.98 for Denmark and Sweden to −1.22 and −1.45 for Australia and the United States. It is plotted against the 1870 real wage in figure 4, and it offers a stunning confirmation of the first hypothesis Between 1870 and 1913 inequality rose dramatically in rich New World countries such as Australia and the United States; inequality fell dramatically in

poor, newly industrializing countries such as Norway, Sweden, Denmark, and Italy; inequality fell only modestly in middle-income, industrial economies such as Belgium, France, Germany, the Netherlands, and the United Kingdom.

The Impact of Globalization on Inequality Trends, 1870–1913

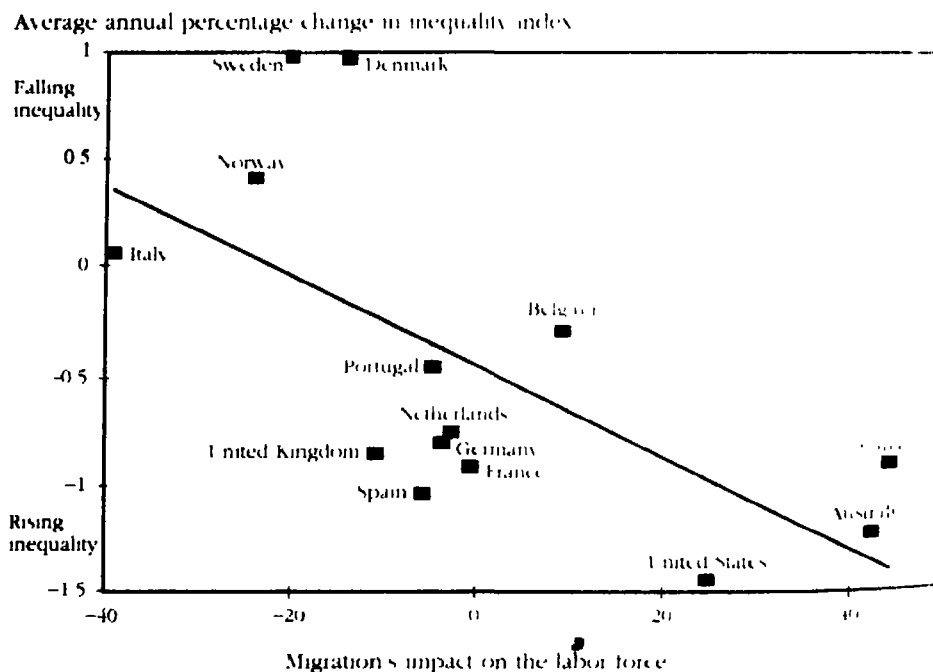
Theory suggests that globalization can account for this key stylized fact: In an age of unrestricted international migration, poor countries should have the highest emigration rates and rich countries should have the highest immigration rates; in an age of liberal trade policy, poor countries should export labor-intensive products and rich countries should import labor-intensive products. Theory is one thing; fact is another. What evidence on trade and migration in the late nineteenth century supports this (apparently plausible) globalization hypothesis?

I start with trade effects. There was a retreat from trade liberalism after 1880, and the retreat included France, Germany, Italy, Portugal, and Spain. In the absence of globalization forces, poor labor-abundant countries that protect do-

Figure 4. *Initial Real Wages vs Inequality Trends, 1870–1913*

mestic industry should raise the returns to scarce factors (such as land) relative to abundant factors (such as unskilled labor). In the face of globalization forces, the same countries should at least mute the rise in the relative scarcity of unskilled labor and thus stem the fall in inequality. The evidence seems to be roughly consistent with these predictions. That is, the correlation between rising inequality and initial labor scarcity turns out to be better for 1870–90—an environment of shared liberal trade polices—than for 1890–1913—an environment of rising protection on the Continent.[9]

I turn next to the impact of mass migration. As indicated above, the impact of mass migration on labor supplies in sending and receiving countries between 1870 and 1910 ranged from 37 percent for three New World destination countries (Canada at 44 percent absorbing the largest supply of immigrant labor) to −18 percent for six poor European sending countries (Italy at −39 percent losing the largest share of its labor supply). Migration's impact on the receiving country's labor force is also known to be highly correlated with an initial scarcity of labor, although not perfectly (Hatton and Williamson 1994). Migration is therefore a prime candidate in accounting for the distribution trends. Figure 5 plots the

Figure 5. *Inequality Trends vs. Migration's Impact on Labor Force, 1870–1913*

Source: See figure 3.

ere immigration increased the receiving country's labor supply, inequality
ɔly; where emigration reduced the sending country's labor supply,
' declined.

unately it is impossible to decompose globalization effects into trade
ition using this information because the correlation between migration's
ıd initial labor scarcity is so high. Yet an effort has been made by
ing a trade-globalization-impact variable as the interaction of initial
city and "openness." The result is that the impact of migration is still
significant, and of the right sign: when immigration rates were small,
an trends were weak; when emigration rates were big, egalitarian trends
ıg; when countries had to accommodate heavy immigration, inegali-
nds were strong. In the Old World periphery, where labor was most
, the more open economies had more egalitarian trends, just as the
r-Ohlin trade model would have predicted. It appears that the open
tigers of that time enjoyed benign egalitarian effects, while those among
ing for autarky did not. In the Old World industrial core, this effect
s powerful. It appears that open economy effects on income distribu-
ambiguous among the land-scarce industrial leaders in Europe where
sector was relatively small.[10] Heckscher and Ohlin would have pre-
s result too In the labor scarce New World, however, the more open
s also had more egalitarian trends, which is certainly *not* what Heckscher
ı would have predicted. The result is not significant, however.

l, I read this evidence as strong support for the impact of mass migra-
ıcome distribution and as weak support for the role of trade. This
exercise explains about two-thirds of the variance in distributional
oss the late nineteenth century What forces could possibly account
naining third, forces that were also highly correlated with initial labor
ıd GDP per worker-hour? Late twentieth century critics of the global-
iesis have argued that the answer lies with technological change.
and Slaughter (1993) contend that a skill-using bias in the United
driven rising inequality. Wood counters that it cannot be so because
in the United States and the other OECD countries was on the rise
the slowdown in productivity was in full swing. Whichever view the
ieves, it is important to remember that we are searching for an expla-
at can account simultaneously for falling inequality in the South, ris-
ality in the North, and some mixture among the newly industrializing
in the middle. But is there any reason to believe that technological
hould be unskilled labor-saving in rich countries and unskilled
ıg in poor countries?

sue has been explored at length (O'Rourke, Taylor, and Williamson
ng the data on the ratio of wages to land rent shown in figure 2.

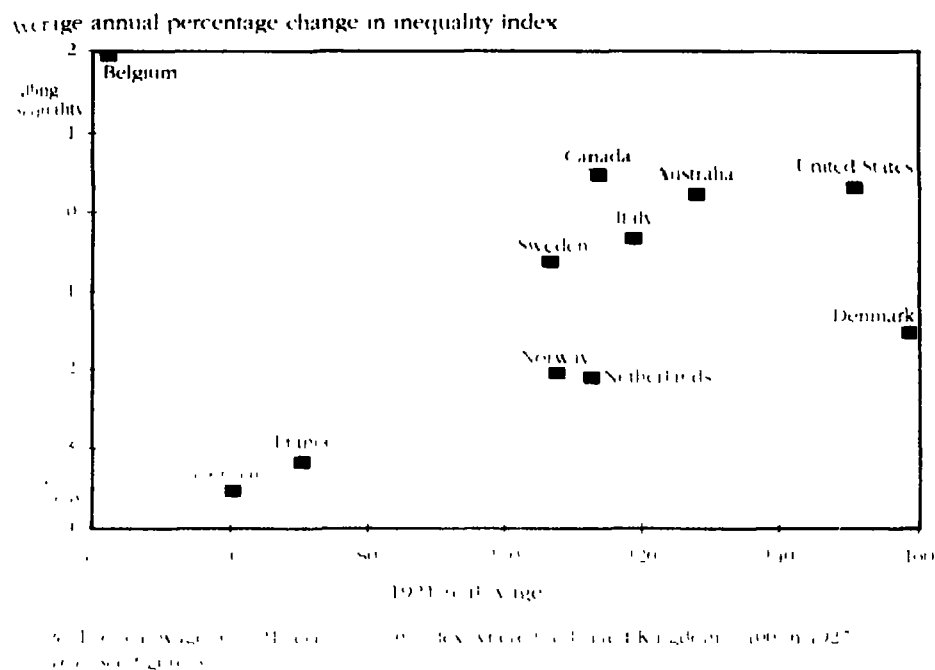
Almost by definition, industrial revolutions embody productivity growth that favors industry. Because industrial output makes little use of farmland, industrialization instead raises the relative demands for labor and capital. Industrial revolutions tend, therefore, to raise wages relative to land rents. According to this prediction, more rapid industrialization in Europe than in the New World should also have raised the wage-rental ratio by more in Europe. Such events should have contributed to a convergence in the prices of factors of production, including a rise in real wages in Europe relative to those in the New World. This prediction would be reinforced if productivity advance in the late nineteenth century New World was labor-saving and land-using, as the above hypothesis suggests and as economic historians generally believe (Habakkuk 1962, David 1974; di Tella 1982). The prediction would be further reinforced if productivity advance in the Old World was land-saving and labor-using, as economic historians generally believe.

O'Rourke, Taylor, and Williamson's results (1996, table 4) are striking. The combination of changes in land-labor ratios and capital deepening accounted for about 26 percent of the fall in the wage-rental ratio in the New World, but for none of its rise in the Old World. Commodity price convergence and Heckscher-Ohlin effects accounted for about 30 percent of the fall in the New World wage-rental ratio and for about 23 percent of its rise in the Old World. Advances in productivity, as predicted, were labor-saving in the labor scarce New World and labor-using in the labor-abundant Old World. Labor-saving technologies appear to have accounted for about 39 percent of the drop in the wage-rental ratio in the New World, while labor-intensive technologies accounted for about 51 percent of its rise in the Old World, powerful technological forces indeed.[11] Globalization accounted for more than half of the rising inequality in rich countries and for a little more than a quarter of the falling inequality in poor ones. Technology accounted for about 40 percent of the rising inequality in rich countries in the forty years before World War I, and about 50 percent of the decline in inequality in poor countries.

Establishing the Inequality Facts, 1921–38

What happened after World War I, when quotas were imposed in immigrating countries, capital markets collapsed, and trade barriers rose?

First, wage differentials between countries widened. Some of the differences were war-related, and some were due to the Depression, but even in the 1920s the trend was clear. Second, the connection between inequality and the forces of globalization was broken (see figure 6). Inequality rose more sharply in poorer countries than in richer countries, where in four cases, it actually declined.

Figure 6. *Initial Real Wage vs Inequality Trends, 1921–1938*

Some Things Never Change

At least two events distinguish the late nineteenth century period of globalization from that of the late twentieth century. First, a decline in inequality seems to have been significant and pervasive in the poor, industrial latecomers in the late nineteenth century sample. This move toward equality has not been universally true of the Latin American and East Asian countries recently studied by other researchers. Second, mass migration appears to have had a more important effect than trade on inequality in the late nineteenth century. Except for the United States, and perhaps West Germany, this phenomenon does not seem to have been true of the late twentieth century, although it should be noted that no economist has assessed the impact of emigration on wages and inequality in Turkey, Mexico, the Philippines, or other developing countries in which net outmigration has been significant over the past quarter century or so.

Some things never change, and that fact implies a warning. Globalization and convergence ceased between 1913 and 1950. It appears that rising inequality in rich countries induced by globalization was responsible, at least in part.

for the interwar retreat from globalization. The connection between globalization and inequality was also broken between World War I and 1950. Rising inequality in the rich countries stopped exactly when immigration was choked off by quotas, global capital markets collapsed, and the international community retreated behind high trade barriers. Are these interwar correlations spurious? The pre–WWI experience suggests not.

Is there a lesson from this history? Will the world economy soon retreat from its commitment to globalization just as it did almost a century ago?

Notes

Jeffrey G. Williamson is the Laird Bell Professor of Economics and faculty affiliate at the Harvard Institute for International Development at Harvard University. A longer and more technical version of this paper appeared as NBER Working Paper 5491 and is available on request (jwilliam@kuznets.fas.harvard.edu). The research underlying this paper was supported by the National Science Foundation. I am grateful for the excellent research assistance of B[illegible] Collins, Asim Khwaja, and Tien Quek, as well as for useful comments from Moe Abramovitz, Don Davis, Ron Findlay, Tim Hatton, Doug Irwin, Ed Leamer, Kevin O'Rourke, Sherman Robinson, Alan Taylor, Adrian Wood, and referees of *The World Bank Research Observer*.

1. See, for example, Baldwin and Cain (1994), Bergstrand and others (1994), Bhagwati and Dehejia (1994), Bhagwati and Kosters (1994), Borjas and Ramey (1994), Freeman (1995), Freeman and Katz (1994), Krugman and Venables (1995), Leamer (1994, 1995), Richardson (1995), Wood (1995a, 1995b), and World Bank (1995).

2. Before 1870 the full sample includes Australia, Belgium, Brazil, France, Germany, Great Britain, Ireland, the Netherlands, Norway, Portugal, Spain, Sweden, and the United States. After 1870, the sample includes Argentina, Canada, Denmark, and Italy.

3. The Smoot-Hawley tariff of 1930 is infamous for its alleged contribution to the Great Depression. Yet the *ad valorem* tariff equivalent levels were 42.5 percent under Smoot-Hawley, an increase of only 8 percentage points over the levels implied by the 1922 Tariff Act (Irwin 1995, table 1). A tariff-induced 8-percentage-point increase seems tiny compared with 45-percentage-point decrease in cost as a result of declining transport charges before World War I—one-sixth the magnitude in fact!

4. Commodity price convergence accounts for about three-tenths of real wage convergence between the United States and Britain during the twenty-five years after 1870 and about one-tenth of the convergence between the United States and Sweden over the four decades after 1870; however, Anglo-American commodity price convergence effects were swamped by other forces after 1895, and they made only a modest contribution to Anglo-Swedish real wage convergence over the four decades as a whole (O'Rourke and Williamson 1994, 1995). O'Rourke, Taylor, and Williamson (1996) turned to econometric analysis of wage-rental trends in seven countries (including Britain and Sweden) to search for the average case. They found that commodity price convergence could explain about a quarter of wage-rental convergence between the New World and the Old World. These estimates are close to the 10–15 percent reported by Richardson (1995, p. 36) for the contribution of trade to rising United States inequality from the 1970s.

5. As far as I am aware, recent studies of the globalizati[illegible]n-inequality connection in developing countries focus almost exclusively on wage inequality, and sometimes only on *urban* wage inequality. I think this is a big mistake for countries where rural wage employment is signifi

cant and where landed interests are powerful. Surely the economic position of landlords and rural labor matters in economies where agriculture is one-fifth, one-quarter, or even one-third of the economy.

6. O'Rourke, Taylor, and Williamson (1996) constructed a panel database documenting the convergence of the ratio of unskilled wages to farm rents per acre among late nineteenth century countries (figure 2): four New World countries—Argentina, Australia, Canada, and the United States; four free-trade Old World countries—Denmark, Great Britain, Ireland, and Sweden; and three protectionist Old World countries—France, Germany, and Spain.

7. This was certainly true of Europe, Argentina, and the American South, but less true for the American Midwest and Canada, where the family farm dominated.

8. The equality index is normalized by setting w/y 1870 = 100.

9. In addition, the slope on an estimated inequality–real-wage regression line is far steeper in 1890–1913 without the protected five (France, Germany, Italy, Portugal, and Spain) than with them. We saw the same contrast when comparing wage-rental ratio trends between four Old World countries with free trade and three Old World countries that are protectionist (figure 2).

10. This was *not* true in England during the 1830s when the Corn Laws can be shown to have had inegalitarian implications and thus that repeal had egalitarian implications (Irwin 1988, Williamson 1990, O'Rourke 1994). England had a very different economic structure and mix of political interest groups in the 1830s compared with the 1880s, a half century later.

11. The residual was 5.1 percent for the New World and 2[illegible]5 percent for the Old.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Baldwin, Robert E., and Glen G. Cain. 1994. "Trade and U.S. Relative Wages: Preliminary Results." Department of Economics, University of Wisconsin, Madison. Processed.

Bergstrand, Jeffrey H., Thomas F. Cosimano, John W. Houck, and Richard G. Sheehan, eds. 1994. *The Changing Distribution of Income in an Open U.S. Economy*. Amsterdam: North Holland.

Bhagwati, Jagdish, and V. Dehejia. 1994. "Free Trade and Wages of the Unskilled: Is Marx Striking Again?" In Jagdish Bhagwati and Marvin H. Kosters, eds., *Trade and Wages: Leveling Wages Down?* Washington, D.C.: AEI Press.

Bhagwati, Jagdish, and Marvin H. Kosters, eds. 1994. *Trade and Wages: Leveling Wages Down?* Washington, D.C.: AEI Press.

Bourguignon, François, and Christian Morrisson. 1991. *External Trade and Income Distribution*. Paris: Organization for Economic and Cooperative Development.

Borjas, George. 1994. "The Economics of Immigration." *Journal of Economic Literature* 32 (December): 1667–1717.

Borjas, George, Richard B. Freeman, and Lawrence F. Katz. 1992. "On the Labor Market Impacts of Immigration and Trade." In George Borjas and Richard B. Freeman, eds., *Immigration and the Work Force: Economic Consequences for the United States and Source Areas*. Chicago: University of Chicago Press.

Borjas, George, and Valerie Ramey. 1994. "Time Series Evidence on the Sources of Trends in Wage Inequality." *American Economic Review* 84(May):10–16.

Burtless, Gary. 1995. "International Trade and the Rise in Earnings Inequality." *Journal of Economic Literature* 33(June):800–16.

David, Paul A. 1974. *Technical Choice, Innovation, and Economic Growth.* Cambridge: Cambridge University Press.

Davis, Donald R. 1996. "Trade Liberalization and Income Distribution." Department of Economics, Harvard University, Cambridge, Mass. Processed.

di Tella, Guido. 1982. "The Economics of the Frontier." In Charles P. Kindleberger and Guido di Tella, eds., *Economics in the Long View,* vol. 1. New York: New York University Press.

Feenstra, Robert C., and Gordon Hanson. 1995. "Foreign Investment, Outsourcing and Relative Wages." In Robert C. Feenstra, Gene M. Grossman, and Douglas A. Irwin, eds., *Political Economy of Trade Policy: Essays in Honor of Jagdish Bhagwati.* Cambridge, Mass.: MIT Press.

Feliciano, Zadia. 1996. "Workers and Trade Liberalization. The Impact of Trade Reforms in Mexico on Wages and Employment." Department of Economics, Queens College, New York. Processed.

Freeman, Richard B. 1995. "Are Your Wages Set in Beijing?" *Journal of Economic Perspectives* 9(Summer):15–32.

———. 1996. *When Earnings Diverge.* New York: Council on Foreign Relations.

Freeman, Richard B., and Lawrence F. Katz. 1994. "Rising Wage Inequality: The United States vs. Other Advanced Countries." In Richard B. Freeman, ed., *Working Under Different Rules.* New York: Russell Sage.

Habakkuk, Hrothgar. J. 1962. *American and British Technology in the Nineteenth Century.* Cambridge: Cambridge University Press.

Hatton, Timothy J., and Jeffrey G. Williamson. 1994. "What Drove the Mass Migrations from Europe in the Late Nineteenth Century?" *Population and Development Review* 20 (September):533–59.

———. 1995. "The Impact of Immigration on American Labor Markets Prior to the Quotas." NBER Working Paper 5185. National Bureau of Economic Research, Cambridge, Mass. Processed.

———. 1997. *The Age of Mass Migration: Causes and Economic Impact.* New York: Oxford University Press.

Irwin, Douglas. A. 1988. "Welfare Effects of British Free Trade: Debate and Evidence from the 1840s." *Journal of Political Economy* 96(December):1142–64.

———. 1995. "From Smoot-Hawley to Reciprocal Trade Agreements: Changing the Course of U.S. Trade Policy in the 1930s." NBER Working Paper 5895. National Bureau of Economic Research, Cambridge, Mass. Processed.

Kosters, Marvin H. 1994. "An Overview of Changing Wage Patterns in the Labor Market." In Jagdish Bhagwati and Marvin H. Kosters, eds., *Trade and Wages: Leveling Wages Down?* Washington, D.C.: AEI Press.

Krueger, Anne. 1978. *Liberalization Attempts and Consequences.* Cambridge, Mass.: Ballinger.

Krugman, Paul, and Anthony Venables. 1995. "Globalization and the Inequality of Nations." NBER Working Paper 5098. National Bureau of Economic Research, Cambridge, Mass. Processed.

Lawrence, Robert Z., and Matthew J. Slaughter. 1993. "International Trade and American Wages in the 1980s: Giant Sucking Sound or Small Hiccup?" *Brookings Papers on Economic Activity, Microeconomics* 2:161–226.

Leamer, Edward E. 1994. "Trade, Wages and Revolving Door Ideas." NBER Working Paper 4716. National Bureau of Economic Research, Cambridge, Mass. Processed.

———. 1995. "A Trade Economist's View of U.S. Wages and Globalization." Anderson Graduate School of Management, UCLA, Los Angeles. Processed.

Maddison, Angus. 1991. *Dynamic Forces in Capitalist Development.* New York: Oxford University Press.

Michaely, Michael, Demetrious Papageorgiu, and Armeane Choksi. 1991. *Liberalizing Foreign Trade: Lessons of Experience in the Developing World,* vol. 7. Cambridge, Mass: Basil Blackwell.

O Rourke, Kevin H. 1994 "The Repeal of the Corn Laws and Irish Emigration." *Explorations in Economic History* 31(1):120–38.

O Rourke, Kevin H., Alan M Taylor, and Jeffrey G Williamson. 1996. "Factor Price Convergence in the Late 19th Century." *International Economic Review* 37(3):499–530.

O Rourke, Kevin H., and Jeffrey G. Williamson. 1994. "Late 19th Century Anglo-American Factor Price Convergence: Were Heckscher and Ohlin Right?" *Journal of Economic History* 54(December):892–916

———. 1995. "Open Economy Forces and Late 19th Century Swedish Catch-Up: A Quantitative Accounting " *Scandinavian Economic History Review* 43(2):171–203

———. 1996. "Education, Globalization, and Catch-Up. Scandinavia in the Swedish Mirror." *Scandinavian Economic History Review* 43(3) 287–309

———. forthcoming "Around the European Periphery 1870–1913: Globalization, Schooling, and Growth." *European Review of Economic History*

Richardson, J David 1995 "Income Inequality and Trade: How to Think, What to Conclude " *Journal of Economic Perspectives* 9(Summer) 33–55

Robbins Donald 1996 "Trade Trade Liberalization, and Inequality in Latin America and East Asia Synthesis of Seven Countries. Harvard Institute for International Development, Cambridge, Mass Processed

Taylor, Alan M., and Jeffrey G. Williamson 1997 "Convergence in the Age of Mass Migration " *European Review of Economic History* 1(April) 27–63

Williamson, Jeffrey G. 1990 "The Impact of Corn Laws Just Prior to Repeal " *Explorations in Economic History* 27(April) 123–56

———. 1995 "The Evolution of Global Labor Markets Since 1830 Background Evidence and Hypotheses " *Explorations in Economic History* 32(April) 141–96

———. 1996a. "Globalization Convergence and History " *Journal of Economic History* 56(June) 277–306

———. 1996b "Globalization and Inequality, Past and Present " Appendixes 1, 2 Harvard University Department of Economics, Cambridge, Mass Processed.

Wood, Adrian 1991 "How Much Does Trade with the South Affect Workers in the North?" *World Bank Research Observer* 6(1) 19–36

———. 1994 *North South Trade, Employment and Inequality Changing Fortunes in a Skill-Driven World* Oxford Clarendon Press

———. 1995a "How Trade Hurt Unskilled Workers " *Journal of Economic Perspectives* 9(Summer):57–80

———. 1995b. "Does Trade Reduce Wage Inequality in Developing Countries?" Institute of Development Studies, University of Sussex, Brighton, U K. July. Processed

World Bank 1995. *World Development Report 1995 Workers in an Integrating World.* New York. Oxford University Press.

The Policymaking Uses of Multitopic Household Survey Data: A Primer

Margaret E. Grosh

Household surveys are a valuable tool for policymakers deciding among policy options. This article illustrates the kinds of issues typically addressed in these surveys and shows how detailed analyses of the data can feed into the decisionmaking process. It outlines the general virtues and limitations of household survey data and provides a brief sample of the kind of data analysis that is relevant to policymaking. It also discusses the benefit-cost ratios that are likely to apply to survey-based policy analysis.

This article is designed to help policymakers and their technical staff understand how household survey data can be useful in their work. Such data may be underused for several other reasons (see box 1), but an appreciation of how useful the analysis of survey data can be in informing the policymaking process increases the likelihood that these other issues will be addressed.

Examples of the Use of Survey Data in Policymaking

Governments can use survey data to inform their policymaking in several different ways, as the following examples illustrate.

Jamaican Food Stamp Reform

In January 1990 the Jamaican cabinet approved major changes in the government food stamp program based, in part, on recommendations arising from an analysis of the Jamaican Survey of Living Conditions (JSLC), the Jamaican version of the World Bank's Living Standards Measurement Study (LSMS). The survey was used at five different stages of the reform process.

The World Bank Research Observer, vol. 12, no. 2 (August 1997), pp. 137–60

Box 1. ***Causes of and Remedies for the Underuse of Survey Data***

Fostering the use of data in policymaking has increasingly become a focus of World Bank attention. Following are descriptions of some of the barriers to the use of data as well as remedies for the specific problem.

Unawareness. Many policymakers are not aware of the many ways that sound data analysis can help formulate policy. As a result many survey projects now include special seminars for policymakers and program managers to brief them on the survey and its potential uses. This article is also aimed at improving awareness.

Restrictions on Access. One of the biggest barriers to data use has been government policies, formal and informal, that deny analysts access to the data collected. The World Bank has attempted with mixed success to promote the availability of LSMS and Social Dimensions of Adjustment survey (SDA) data in the countries in which they were collected. Grosh and Munoz (1996) describe policy and logistical requirements for ensuring widespread access to data; Hartke (forthcoming) offers guidance on how statistical agencies can disseminate data systematically.

The World Bank maintains central data archives for both LSMS and SDA. Grosh and Glewwe (1995) and the LSMS Home Page (http://www.worldbank.org/html/prdph/lsms/lsmshome.html) describe the data available through the LSMS archives and the procedures for obtaining the data. The "catalogue" and web site are the culmination of a multiyear effort to improve the documentation and dissemination of the LSMS surveys. Many of the survey datasets are now available to the public without restriction. Some require government permission, which is usually given promptly.

Scarce Analytical Capacity. Projects to support data collection increasingly also support the creation of analytic capacity, through training, provision of hardware and software, studies, and funding. Papers by Blank and Grosh (1996) and Synergie (1996) synthesize lessons from attempts to build analytic capacity in conjunction with LSMS and SDA surveys, respectively.

Inadequate Communication. In many projects, the survey is guided by a formal steering committee composed of policymakers and academics so that data users and data collectors have a forum for discussion. Grosh (1991) covers some of the institutional issues involved in ensuring that household survey data have an effect on policy decisions.

Stage 1. The first survey was conducted in August 1988. The preliminary abstract produced two months later showed that the food stamp program was better targeted to people in need than were food subsidies in general (STATIN and World Bank 1988). Thirty-one percent of the food stamps distributed went to the poorest quintile, while only 14 percent of the general food subsidies reached this target population. Although the overall finding was not unexpected the fact that the size of the difference could be quantified probably added support to arguments in favor of reducing food subsidies and increasing the budget for food stamps (Grosh 1995).

Stage 2. In July 1989 a second round of the survey was fielded. An abstract drafted two months later showed that malnourished children used the public health system as often as those who were not malnourished, and that children who did not receive food stamps used the public clinics almost as often as those who did (STATIN and PIOJ 1989). Thus the food stamps directed to children

ırough clinics could continue to be delivered through health clinics without ndue concern that the neediest children would be missed.

Stage 3. The Planning Institute of Jamaica commissioned a study of the surey data to help determine whether it was necessary to change the benefit levels nd the criteria defining eligibility for the food stamp program (Gordon 1989; nderson 1989). That report was completed in December 1989, and the new enefit levels and eligibility criteria, which largely followed the recommendaons, were approved in January 1990.

Stage 4. To determine what changes in administrative procedures were likely enhance the coverage of children and pregnant and lactating women, the Planıng Institute commissioned further work, using existing survey data and the lministrative records of the food stamp program (Anderson 1993). A special ırvey of mothers at urban health clinics was also carried out, along with observaons at food stamp pay stations and interviews with program employees.

Stage 5. A special section on food stamps was added to the survey in 1989 nd revised each year until 1992 and periodically thereafter. This section was esigned to address various issues raised by policy analysts. Basic tabulations on ow many people benefit, where they fall in the distribution of welfare, and hy some people do not apply for food stamps have been reported annually in ie Jamaican Survey of Living Conditions abstracts. The data collected were sed to complement the administrative records used in the routine monitoring f the program.

The analyses were timely, well-suited to the policy question, clearly explained the policy audience, and carried out in a logical, sequential fashion, largely in sponse to queries from policymakers at each stage of the process. Much of the ork was also peer-reviewed and was made available to the public. Governnent officials, academics, and World Bank staff were all involved in the analys. The questions on the SLC relating to the food stamp program were changed veral times to make them more relevant to policy issues.

The Jamaican example is noteworthy not only because it demonstrates the any advantages of using survey data in policymaking, but also because such ear illustrations are rare. Other examples are less clearcut, although just as nportant. Two other examples of the use of survey data in policymaking are ınder stories to tell.

Ghana and Peru

n 1990 the government of Ghana was considering an increase in the tax on etroleum products but was eager to cushion the impact of the price increase n the poor. Officials first considered subsidizing the price of kerosene to mitiate the effect of the tax increase. Data from the Ghana Living Standards Sur-

vey showed that the poor did not consume much kerosene, however, but instead used wood as fuel for cooking. Thus much of the subsidy would have benefited the nonpoor rather than the poor. The survey data prevented the government from making an inefficient and costly policy decision. Yet no special analysis was commissioned, and no conference was held; a few people associated with the policymaking process looked at some simple tables that were already available and made a better policy decision.

A dramatic contrast to Ghana's low-key use of household survey data is the story of welfare analysis in Peru using the LSMS surveys conducted in 1985, 1990, 1991, and 1994. The analysis is very visible, but its impact is more difficult to pin down.

The first survey was conducted by the government of Peru; subsequent surveys were carried out by a private firm. The results of the surveys are available in bookstores in Peru, are often quoted in newspapers and in widely read publications on national issues, and are easily available to academic researchers. The data have been used in policymaking (for example, an analysis of these statistics informed the poverty alleviation strategy that the government presented to a consortium of development agencies in 1993). The surveys showed that poverty had increased sharply during the economic crisis from 1985 to 1990 but that it had declined somewhat after macroeconomic reforms were introduced in 1991. A widely known, technically respected, and factual basis for statements about the level of poverty or the effect of macroeconomic policies on the poor must surely affect the way in which the debate over appropriate policies takes place either in the proverbial smoke-filled back rooms or in the streets. The survey's influence on the political debate cannot be measured, however.

These three examples show that there is no single formula for using household survey data to help make public policy. In all cases, however, two elements are necessary: the data must be analyzed in a manner pertinent to the policy question, and the analysis must be transmitted to policymakers. In the cases cited here, the analysts were sufficiently well-informed about the agenda to produce policy-relevant analysis, and they had the means to bring their results to the attention of policymakers who, fortunately, were receptive to the analysis. In Jamaica and Peru, policymakers actually determined the direction of the analysis and commissioned work to answer the relevant questions.

The Benefits and Limitations of National Survey Data

Before policymakers commission analyses of survey data, it is important for them to appreciate the benefits and limitations of such data. The discussion

here is based on surveys that are similar to LSMS and SDA surveys, that is, they are multitopic surveys designed to study household welfare, household behavior, and the effects of government policies. (See box 2 for an outline of the key features of these surveys.) Most of the lessons learned from these exercises also apply to data from other kinds of multitopic household surveys, which offer a broader range of interesting applications than do single-topic surveys.

Benefits

The first benefit of a good household survey is that it provides hard information, enabling policymakers to move from polemics to concrete and constructive discussions. For example, in 1993 the World Bank and the government of Guyana met to discuss health care financing. The Ministry of Health was not very interested in charging user fees for public health services. It saw such fees as part of an ideology that the Guyanese government did not share and that was not conducive to providing health care to the poor. But the 1993 Guyanese Survey of Living Conditions showed that only 61 percent of those in the poorest quintile who sought health care used the public sector (World Bank 1994, table 4.5). This fact helped to establish common ground in the dialogue. Because many of the poor were already paying for health care, the theoretical discussion of whether user fees should exist was rendered moot, and thus it was possible to proceed with a more practical discussion of how to improve the health of the poor with a mix of already existing public and private services.

A second benefit is that survey data help to answer the "what if" question in assessing the impact of actual or proposed policy changes. For example, the governments of Jamaica and Indonesia have both used household surveys to estimate how changes in user fees for public health services would affect the use of public health facilities and the revenues collected.

The third benefit, and one that is unique to surveys, is that they represent the whole population, including the rich and the poor, those who use public services and those who do not. In contrast, statistics gathered in the course of delivering public services are inherently biased because they do not include the population that is not receiving the service. Often those are the people about whom policymakers have most reason to be concerned; examples are unvaccinated children or those who are not in school. To illustrate, consider the differences in monitoring children's nutritional status using clinic-based, rather than survey, data. The children at clinics may not be a random sample of all children and may be sicker, poorer, and more malnourished. Alternatively, children who use clinics frequently may receive adequate preventive care and thus be less likely to be sick or malnourished than other children. Poor children may not use clinics because they are too distant. Conversely, wealthier children may

Box 2. ***The Characteristics of Nationally Representative Multitopic Surveys***

The World Bank's Living Standards Measurement Study (LSMS) and Social Dimensions of Adjustment (SDA) surveys are programs designed to help governments gather comprehensive information on household welfare and to promote analyses of the data. Since 1985 such surveys, which attempt to find out how households behave and are affected by economic policies, have been implemented in some fifty countries (for a full explanation of these surveys, see Grosh and Glewwe 1995; Simonpietri and Ngong 1995).

Extensive Multitopic Questionnaires

The LSMS and SDA Integrated Survey questionnaires, which are designed to cover individual and household-specific information, are often complemented by separate questionnaires on prices and community services. The kind and extent of information gathered—and therefore the questionnaire used—vary greatly from country to country, but the surveys typically include the following:

- *Consumption.* Explicit food and nonfood expenditures, the value of home-produced food and food received as gifts, information to impute the value of durable goods and owner occupied housing
- *Income.* Wages and in-kind benefits from employment, net revenue from farming and from nonagricultural household enterprises.
- *Social dimensions of welfare.* Nutritional and health status, literacy, availability of water quality of housing.
- *Access to and use of public services and subsidies* Schools, clinics, agricultural extension services, welfare programs.
- *Other household decisions* Fertility, migration, labor force participation
- *Local conditions.* Prices, labor markets, availability of transport, commercial services

The SDA Priority Surveys, which were designed for monitoring trends rather than for providing data for analyses, usually cover a subset of the indicators above They often use shortcut measur

receive most of their health care from private physicians, who are not included in the nutrition reporting system. The direction and extent of the bias in clinic-based data is not clear. (See Grosh, Fox, and Jackson 1991 for a more complete discussion of this issue and a review of empirical evidence.)

Moreover, this bias may not be constant over time, making trends drawn from clinic-based data unreliable. During an economic crisis, for example, wages and employment fall and prices rise, which is likely to increase the incidence of malnutrition. And the pattern of use at clinics may change at the same time Some families may be forced out of the system altogether, while others may forgo preventive visits but still use the clinics when their children are ill. Those who previously used private physicians may begin using public care. These kinds of changes in patterns of health care use would affect the amount of bias in the malnutrition rate as measured from clinic data, making any inferences about the effect of the economic crisis on malnutrition questionable. A carefully drawn household survey could elminate these biases.

of consumption, omit income measures altogether, and collect a reduced set of all others types of information.

Quality Assurance Mechanisms

The LSMS and SDA survey programs rely on a variety of mechanisms to preserve data quality (see Grosh and Muñoz 1996, for the LSMS surveys; Delaine and others 1992 for the SDA surveys). Among the most important are:

- *Questionnaire design.* In-depth consultation with policymakers and analysts; extensive field testing.
- *Questionnaire format.* Precoding, explicit wording, filtered questions, instructions to interviewers, use of graphic design to minimize interviewer error.
- *Training and supervising field teams.* Maintaining high supervisor-to-interviewer ratios, often 1 to 2.
- *Concurrent data entry.* Ensuring that quality control features are embedded in the data entry program so that mistakes encountered when the data are being entered can be corrected by revisiting the household.

Small Nationally Representative Samples

Most multitopic surveys that share the characteristics mentioned here aim to cover an entire country. The LSMS and SDA Integrated Surveys are national samples covering from 2,000 to 5,000 households, to ensure high-quality fieldwork (and control for nonsampling error), to use statistical rigor. In budgetary terms, the relatively small size of the sample helps to counterbalance the costs of such extensive questionnaires and of quality control. Because SDA Priority Surveys are simpler, they can cover 8,000 to 10,000 households.

A fourth benefit is that surveys provide a reference group to which data from other special focus or special sample surveys can be compared. This can work in several ways. In Bolivia a survey of the labor force was used as a comparison group for a special survey of workers on the labor-intensive public works projects funded by the Emergency Social Fund. Unfortunately, the labor force survey was limited to urban areas, which meant that the evaluation of the impact of the public works projects had to be limited to urban areas.

And finally survey data can be used to study the links among many different topics, such as the effects of education on earnings or fertility, or the effects of health status on labor force participation or productivity, or the effects of welfare transfers on consumption behavior. To support such intersectoral analysis, however, surveys must gather information on many topics, which implies that much less information will be included on any one topic than would typically be gathered in a single-purpose survey. So this virtue carries with it a limitation in some respects.

This point is illustrated by a comparison of the information on fertility gathered through the LSMS with that obtained through the Demographic and Health Surveys, the state-of-the-art single-topic surveys on demography funded by the U.S. Agency for International Development and conducted by a private agency. The LSMS usually collects fertility information on a narrow range of demographic factors that basically relate to a woman's pregnancy and birth history and to her use of maternity services. Demographic and Health Surveys also measure these factors, but in addition they provide information on the woman's contraceptive knowledge and use, marriage history, and the number of children desired. Moreover, the samples in the single-topic surveys represent women aged fifteen through forty-nine, rather than the whole population, and are usually somewhat larger than the samples used for multitopic surveys. The single topic surveys, however, contain very little information on the woman's income, consumption, labor activities, education, or health. Thus, while single-topic surveys are more accurate in measuring a wide spectrum of demographic variables related to fertility, multitopic surveys are better for studying the factors that determine fertility.

Limitations

Household surveys, even when they are regularly conducted, cannot address all policy questions. It is important to bear in mind three inherent limitations of such data. The first limitation is imposed by the size of the sample. The fairly small national samples do not allow reliable study of "rare events." Examples of rare events that are of interest to policymakers are infant mortality and enrollment in small government programs. Infant mortality by definition affects only children under twelve months, who usually account for about 2 percent of the population. Even with high infant mortality rates of 100 per 1,000 live births or one-tenth of the cohort, only 0.2 percent of the sample would be affected. Thus, in a sample of 2,000 households (or 10,000 individuals), only 20 infant deaths would be detected. This is too low a figure to produce precise estimates of levels of infant mortality and certainly too low to study patterns.

This problem can be partly mitigated by asking about deaths over a longer period, say, five years rather than one, so that 100 deaths might be observed. Alternatively, the sample could be doubled to 4,000 households. Nonetheless, some events, although important, are not frequent enough to study with a small general sample. The planners of an LSMS conducted in Nicaragua in 1993, for example, were interested in studying the welfare of individuals disabled during the civil war. In the sample of 3,600 households, however, only eight people reported that their mobility was limited because of war wounds, and only fifteen people reported having war-related deformities.

The other aspect of the limitation imposed by the sample size is that the data cannot be reliably disaggregated into small subgroups, even for events that are not rare. Consider, for example, the issue of access to water. Policymakers commonly want statistics to apply to small geographic areas such as provinces or districts, especially where these levels of government are in charge of the infrastructure for water supply. A country could easily have twenty provinces, however, which would mean that for a survey of 2,000 households, the sample would include an average of only 100 families in each province—and fewer in smaller provinces. Moreover, it would be useful to contrast, for example, rural or urban households within the province, or poor and nonpoor, which would further shrink the number of observations from each group. The precision of the resulting estimates would be very low, and would entirely preclude the possibility of disaggregating the figures to lower levels, such as by district.

The second limitation stems from the survey's reliance on formal interviews that use predetermined, closed-end questions. Such questions are conducive to gathering some kinds of factual information but are less suited to studying perceptions, motivations, or nuances of opinion. In a few cases, notably the South Africa LSMS survey but also the 1993 Jamaica Survey of Living Conditions discussed earlier, a few closed-end attitudinal questions were included. Such questions could be added more regularly, but they can barely scratch the surface of such issues. To study fully attitudes, perceptions, or motivations, open-ended interviews or focus groups are more appropriate.

Moreover, the predetermined nature of formal questionnaires means that the survey planners have to know in advance what to measure, which in turn implies that they already know a lot about the phenomenon they want to study. To gather that knowledge, the survey planners can use qualitative studies as a foundation for learning which factors are likely to influence the outcomes that are being measured. Then the survey questionnaire can be better designed.

In Jamaica findings based on data from the Jamaican Survey of Living Conditions (World Bank 1995b) were compared with those from a participatory urban appraisal conducted in 1995 (Moser and Holland 1995). Many of the findings of these two different kinds of studies were broadly consistent—that the poor need better jobs, more human capital, and improved physical infrastructure and that crime and violence are significant obstacles to achieving these goals. The poverty assessment (which was *quantitatively* based) concluded broadly that growth and investment in basic services and infrastructure would aid the poor. The participatory urban appraisal (which was *qualitatively* based) was able to add subtleties to this basic understanding. For instance, the poor perceived the existence of an "area stigma" with respect to jobs. In other words, potential employers did not hire poor applicants after they revealed where they

lived. This information indicated that macroeconomic growth alone would not provide jobs for the poor.

Similarly, the 1989 survey questionnaire in Jamaica included a set of questions on the distance from the respondent's household to various public services. The results showed that these distances were generally not great, which was interpreted as meaning that the poor had access to these services. The participatory urban appraisal, however, revealed that violence in some urban areas often restricted residents' access to nearby services because there were only a limited number of hours in the day when people felt safe to move about or because they could not cross "turf lines" without encountering gang activity. Again, the qualitative work was able to identify problems (and thereby lead to solutions) that had not been observed in the formal, closed-end interviews conducted in the course of the quantitative survey.

A third limitation is that household survey data may not be sufficient to answer many important policy questions and thus may need to be supplemented with data from other sources. For example, at least four types of data are needed to find out how schools could teach students most efficiently. The first kind of data—the characteristics of the students and their families, such as age, education level, income, and work activities—can easily be gathered in a household survey. The second—information about students' scholastic levels—requires achievement tests. Such tests can be organized as part of a household survey although it is much more common (and many educators would say, more reliable) to use a school-based survey.[1] Third, researchers need to obtain data on the educational process—the curriculum; teachers' qualifications; provision of supplies and equipment such as textbooks, blackboards, reference books, classrooms, and lab equipment—as well as data on such factors as how management affects these inputs. These data must be collected in schools. And finally, it is important to know the costs of providing materials and programs that affect how teachers teach. This data is most likely to be available from education planning offices.

Although this complementary data can be added to a household survey (as it was in 1988 in Ghana, in 1990 in Jamaica, and in 1990–91 in Morocco), the exercise is not easy. Developing the necessary instruments adds significantly to the complexities of planning, especially if it is necessary to identify or develop tests that produce valid comparisons across, for example, children in a wide range of grades (which is usually the case because the sample households do not have many children in each grade). Moreover, determining the sample of schools that should be included is difficult, particularly in urban areas where children have a choice among different schools. And finally, administering the tests and school questionnaires requires more fieldwork, although this is by far the most easily handled aspect of collecting the required ancillary data.

What Topics Are Suitable for Policy Analysis?

The range of potential policies that can be affected by survey data is very broad. All of the examples discussed here have at least been disseminated in policymaking circles, and in most cases the government agency involved actively participated in crafting the data collection and analyses, and the analyses were taken into account in the policymaking process. Space constrains the number of examples presented here, but I have tried to show the diversity available both in terms of the analytic issue that can be addressed and the sector to which the analysis pertains.

Policymakers need to consider four areas in connection with the issue to be analyzed. First, what are the outcomes that are of interest, such as the employment rate or the percentage of children who are malnourished? Second, what are the patterns in the use of the service? Who uses health clinics? Schools? Public transportation? Third, how will changes in policies (higher fees, say) or in the economic environment affect the community or the providers of the service? How does economic growth affect the poorest? And finally, what determines household behavior pertaining to the issue being analyzed, such as the number of children a family will have, or whether the children attend school? The range of sectors in which these different questions can be addressed is as broad as the set of topics in the questionnaires: health, fertility, nutrition, education, migration, employment, agriculture, housing, consumption, small business, and ownership of assets. The following examples illustrate the point.

The Study of Poverty

One way to evaluate the prevalence of poverty is to construct a poverty profile. Poverty profiles quantify and describe several dimensions of poverty, including who the poor are, where they live, how they earn their living, whether they have access to and use of government services and subsidies, and what their standard of living is with regard to health, education, nutrition, and so forth. To cover these aspects of poverty, researchers use information from many parts of a multitopic questionnaire. Here I present part of a single table from the Ecuador Poverty Report (see table 1); the full report has many such tables.

Another way that surveys help to study poverty is to show how it changes over time. In the late 1980s Peru's economy experienced considerable upheaval. Gross domestic product per capita fell by about a fourth. The price index rose from 3,474 in 1985 to 40,216,592 in 1990. Net international reserves plummeted. Using data from 1985 and 1990 household surveys, Glewwe and Hall (1994) found that the consumption of the average household in Lima fell by slightly more than half during this period and that the welfare of the poorest

Table 1. *Ecuador: Some Characteristics of the Poor, 1994*

Characteristic	Region	Urban Poor	Urban Nonpoor	Rural Poor	Rural Nonpoor	Total Poor	Total Nonpoor
Education							
Education of household head (years)	National	5.2	9.1	3.2	4.7	4.0	7.5
	Costa	4.9	8.3	2.8	3.9	3.9	7.1
	Sierra	5.8	10.5	3.4	5.1	4.1	8.0
	Oriente	5.9	8.8	4.5	7.4	4.6	7.8
Health							
Diseases treated informally (percent)	National	24.8	14.8	32.7	24.1	29.4	18.0
	Costa	27.3	19.0	45.3	33.7	36.4	22.6
	Sierra	19.7	9.6	21.4	19.4	20.8	13.7
	Oriente	26.3	10.7	20.1	14.4	20.4	13.2
Employment							
Informal sector (percent)	National	54.6	44.1	27.9	35.8	39.2	41.7
	Costa	54.6	44.1	19.6	24.8	37.6	41.6
	Sierra	56.3	41.3	35.1	42.6	42.3	41.9
	Oriente	54.9	40.8	25.7	41.1	27.3	40.9
Regulated sector (percent)	National	15.5	35.3	3.4	9.9	8.6	26.7
	Costa	11.8	31.1	1.1	3.1	6.6	24.4
	Sierra	22.1	41.3	5.4	12.6	11.1	29.2
	Oriente	8.7	40.0	6.4	26.8	6.5	31.0
Basic services							
Sewerage connection (percent)	National	57.3	83.4	12.4	28.2	29.6	[illegible]
	Costa	43.5	74.4	11.7	17.0	27.3	58.9
	Sierra	78.9	95.6	13.5	35.4	33.5	69.8
	Oriente	62.9	87.9	7.0	31.1	10.8	50.6
Electricity supply (percent)	National	97.8	99.5	62.0	75.8	75.8	91.1
	Costa	97.9	99.4	55.5	63.3	76.4	89.6
	Sierra	97.7	99.7	69.8	84.3	78.4	93.0
	Oriente	93.6	96.5	36.3	74.4	40.1	81.9
Water from public network (percent)	National	61.2	78.8	18.3	23.0	34.8	[illegible]
	Costa	48.9	67.1	6.1	9.1	27.2	[illegible]
	Sierra	79.9	94.5	27.9	34.0	43.8	68.2
	Oriente	85.3	92.5	12.1	23.2	17.0	[illegible]
Waste collection (percent)	National	59.7	76.7	1.1	5.6	23.5	51.5
	Costal	52.2	68.9	1.3	6.8	26.6	[illegible]
	Sierra	70.5	87.7	0.9	3.9	22.2	51.5
	Oriente	59.9	84.9	1.8	21.5	5.7	43.3

Source: World Bank (1995a), tables 2a and 2b

Table 2. *Lima: Changes in Consumption Expenditures, 1985 to 1990*

Characteristics of heads of households	*Percentage change in expenditures*
Gender	
Male	−54.5
Female	−54.9
Education	
None	−58.7
Primary	−59.1
Secondary general	−55.1
University	−54.0
Other postsecondary	−38.9
Employer	
Government	56.1
Private	−56.3
Private home	−54.7
Self-employed	51.4
Occupation	
Agriculture	50.3
Sales/services	−56.7
Industry/crafts	52.3
White collar	−54.1
Unemployed	−65.9
Retired	50.2
All Lima	54.6

Source: Glewwe and Hall 1994.

households dropped even more than the average (table 2). Households headed by individuals with little or no education experienced the greatest loss of welfare. Female-headed households did not fare worse than households headed by men. Poverty, defined as the inability to cover a household's basic nutritional requirements, increased from 0.5 percent of the population to 17.3 percent.

Understanding the Effects of the Economic Environment

Household survey data can provide information on the effects of changes in taxes, subsidies, or trade policies on individuals or groups. The following example from Tunisia illustrates an attempt to assess the impact of a change in the price of a consumer good.[2] Analogous work can be done for price changes on goods produced by farming households (see, for example, Deaton and Benjamin 1993).

For many years the Tunisian government subsidized the consumer prices of several staple goods. After 1990 the government began to change incrementally

the amount of the subsidy and the commodities included in the program in an attempt to increase the effectiveness and reduce the costs of the subsidy program. Table 3 shows some analysis done in the course of discussions between the government of Tunisia and the World Bank (see Tuck and Lindert 1996) to determine what policy changes should be adopted. The effect of various price changes on households' caloric intake by expenditure quintile was simulated, taking into account changes in the consumption of specific foodstuffs as a result of price changes, holding all other factors constant. The simulation was based on data from a survey conducted especially to help guide decisions about subsidy reforms. Analysts estimated that a 50 percent reduction in subsidies across the board would reduce the caloric intake of the poorest quintile by 30 percent. Targeted cuts in the subsidies of specific goods, however, were expected to lead to a much smaller reduction (about 19 percent), although simulations revealed that both scenarios would generate comparable fiscal savings for the government. Not surprisingly, the government adopted a strategy that included targeted changes in subsidies.

Table 3. *Tunisia: Estimated Nutritional Effects of Alternative Price Policies*

Impact of hypothetical price changes	*Expenditure quintile* 1 (Poorest)	2	3	4	5 (Richest)	Average
(1) Subsidy cut of 50 percent						
Percentage change in calories as share of total caloric intake	−30.1	−24.3	−22.2	−20.6	−15.3	[illegible]
Resulting caloric intake	1,483	1,688	1,813	1,975	2,549	[illegible]
(2) Targeted cut						
Percentage change in calories as share of total caloric intake	−19.5	−20.9	−22.6	22.6	22.5	[illegible]
Resulting caloric intake	1,708	1,764	1,803	1,925	2,332	[illegible]
Base case						
1993 levels (Kcal)	2,122	2,230	2,330	2,487	3,009	[illegible]
Subsidized goods as a share of total intake (1993)	58.9	49.4	47.4	42.4	28.4	[illegible]

Note: Scenario (1): Impact of cutting subsidies by 50 percent from 1993 levels on quantities consumed. Scenario (2): Impact of eliminating subsidies on specific goods on quantities consumed (sterilized milk, *gros pain*, bottled generic oil). A negative number signals a loss in caloric intake. Estimates omit introduction of new goods since 1993. Recommended daily allowance: 2,165 calories per capita (National Statistical Institute).

Source: Tuck and Lindert (1996), tables 27 and 28.

The Provision of Public Services

The first question to address in thinking about service provision is who has access to these services. Findings from a subset of the information available for rural areas from the Viet Nam household survey (World Bank 1995c) showed that the poor had less access to services than the nonpoor but that the differences were relatively small (table 4). Health facilities are more accessible generally in the south than they are in the north, but the reverse is true of agricultural services and literacy programs.

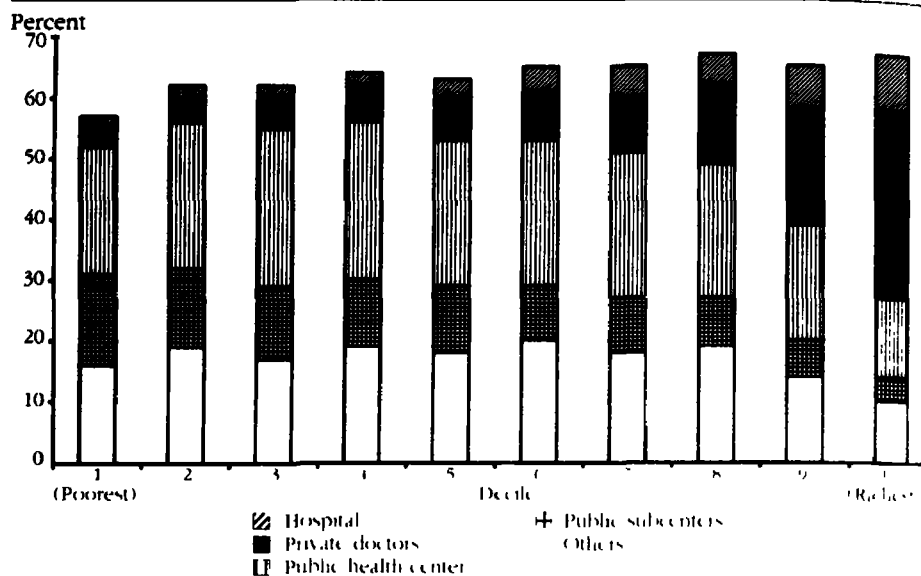
The second question is who uses public services. Household surveys that include appropriate questions can answer this question. Figure 1 shows some results from a 1990 survey in Indonesia. Among respondents who were ill during the month preceding the fieldwork, 33 percent of those in the richest decile did not seek health care, compared with 44 percent of those in the poorest

Table 4. *Viet Nam Percent of Population Living in Rural Communes with Access to Infrastructure*

Infrastructure	South			North		
	Total	*Nonpoor*	*Poor*	*Total*	*Nonpoor*	*Poor*
Passable road	58.0	58.1	57.9	76.8	88.5	69.4
Public transport	61.2	61.1	61.3	47.2	54.3	42.7
Electricity/generator	91.6	91.6	91.6	85.6	90.0	82.8
Pipe borne water	7.5	9.3	5.8	3.6	5.6	2.3
Permanent market	71.5	72.6	70.4	43.5	55.6	35.8
Post office	46.8	43.4	50.3	27.7	28.9	26.9
Lower secondary school	82.9	81.[illegible]	83.8	90.6	92.6	84.9
Upper secondary school	10.6	12.3	8.9	9.3	9.4	9.3
Dispensary	55.6	60.0	51.3	19.7	20.0	19.6
Pharmacy	8.3	80.7	76.0	65.5	72.0	61.3
[illegible]	92.2	90.1	94.2	93.9	97.1	91.9
Doctor	50.9	60.8	41.0	34.7	42.5	29.8
[illegible]	100.0	100.0	100.0	94.0	96.8	92.2
[illegible]	94.4	95.2	93.7	88.4	88.8	88.2
Agricultural extension office	18.4	22.2	14.5	27.8	29.9	26.4
Agricultural extension agent visits	72.1	68.9	75.3	71.3	75.8	68.3
Cooperative	8.7	8.9	8.4	90.6	94.2	88.3
Adult literacy program	81.9	81.0	82.8	85.3	86.9	84.3
[illegible] exchange	93.0	92.7	93.4	97.4	97.1	97.6

Note: The poverty line used is calculated for seven different regions and separately for urban and rural areas in each region. The national average poverty line is 1,117 thousand dong per person per year.

Source: World Bank (1995c), annex 3.1, tables 4 and 5, pp. 170–71.

Figure 1. *Indonesia: Percentage of Those Ill in Past Month Who Sought Health Care, by Decile and Place Where Care Sought*

Note: According to 1990 National Economic Social Survey.
Source: World Bank (1993) figure 1.10, p. 18.

decile. Of those in the poorest decile, 37 percent went to public health centers while only 3 percent sought care from private physicians. In contrast, in the richest decile, only 17 percent used public health centers and 31 percent relied on private doctors.

The third question concerns how the value of the subsidy is distributed. To answer this question, information on the use of services from the household survey must be supplemented with information on the costs of providing services. This figure can come either from budget accounts or from special studies. When such information is available, it is possible to conduct analyses like that shown for Indonesia in table 5. The value of subsidies to education is greater than the combined value of subsidies to health and to consumption of kerosene. The absolute value of the subsidy captured by the richest decile is two to four times greater than the absolute value of the subsidy captured by the poorest decile. The share of household expenditure accounted for by the subsidies is greater for the poor than the rich, however, indicating that these factors do help to equalize the distribution of welfare.

A final concern is what would happen if user fees were raised. An important policy question in several sectors is whether charging (or increasing) user fees would affect the use of services and the revenues of the service providers. Using

Table 5. *Indonesia: Distribution of Selected Subsidies*

Subsidy	*Year*	*Decile* *Poorest*	*Decile* *Richest*	*National average*
Per capita (rupees per month)				
Education	1989	1,161	2,469	1,520
Health	1989	113	313	213
Kerosene	1990	94	447	243
Percentage of household expenditure				
Education	1989	13.18	4.04	6.57
Health	1989	1.00	0.38	0.70
Kerosene	1990	0.84	0.56	0.82

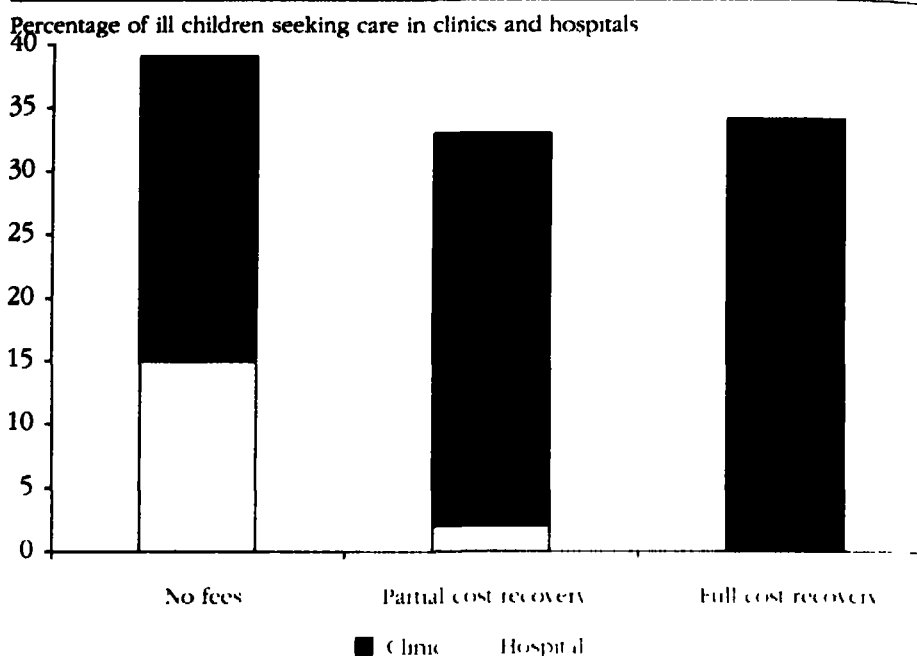
Source: World Bank (1993), annex 2.2, tables 3, 4, 8, 9, 13, 14.

LSMS data, researchers have analyzed both of these factors, primarily in the health sector but also in education. Figure 2 presents a simulation by Gertler and van der Gaag (1990) showing how the use of health services for children in rural areas of Côte d'Ivoire might change in response to alternative pricing policies. Using econometric techniques, Gertler and van der Gaag attempted to compare the number of children that would seek care at (fee-charging) hospitals and (free) clinics with the number that were presently treated under a no charge policy. They found that the introduction of such fees in hospitals slightly reduced the percentage of sick children seeking health care (from 38 percent under a no-charge policy to 33 percent with user fees), but that virtually all those seeking care would switch from hospitals to free clinics.

The Impact of Government Programs

Finally, household survey data can be used to assess the effect of government programs. Impact evaluations often require special sampling or other data sets to complement household survey data, but in the cases that follow, the special design features were kept fairly simple.

In the first case, Bolivian officials wanted to evaluate the effect of a public works program on poor workers. Such programs are often part of the effort to alleviate poverty. The idea is that the jobs will be self-targeted because only the truly poor are willing to accept temporary jobs that require hard physical labor and pay low wages. It is important not only to evaluate whether this self-targeting is effective but also how much workers benefit. If the public works program did not exist, poor workers who could not afford to be completely idle might instead be selling chewing gum on street corners or turning up each day at places where daily laborers are hired. The earnings from these other activities might be low, but they would bring in some income. Thus, for the workers, the monetary ben-

Figure 2. ***Côte d'Ivoire: Simulated Impact of Raising Hospital Fees in Rural Areas***

Note: The source did not give the tables with the exact numbers. These graphs are approximations of the originals based on visual inspection.

Source: Gertler and van der Gaag (1990), figure 7.2, p. 106.

efit of a public works job is the difference between the wage it pays and whatever the workers might be able to earn in their alternative activities.

Thus, to evaluate the benefits from the public works programs financed by the Bolivian Emergency Social Fund, officials added a supplementary sample to the 1988 Permanent Survey, which was carried out periodically in urban areas throughout the country. In this supplementary sample, laborers on public works projects were interviewed using a questionnaire that included the questions from the 1988 Permanent Survey plus some additional questions. The two data sets were combined to estimate what the workers would have earned had they not been employed by the Emergency Social Fund. Newman, Jorgensen, and Pradhan (1992) analyzed the data and found that without the job program, 77 percent of workers would have been in the bottom four income deciles. Thus, the program was well-targeted. Moreover, it raised the distribution of income; earnings of workers increased over their preprogram level by 45 percent a week.

The second case involves the effect of government transfers on private transfers. Private, nonmarket transfers (such as remittances from abroad and support from the family) occur almost everywhere in the world, but they are an especially important part of economic life in developing countries. While 15 percent of individuals in the United States report receiving transfers, the figure in developing countries is 19–47 percent (Cox and Jimenez 1993). Thus the appropriate size of the public safety net depends in part on the size of the private safety net that is already in place.

Household surveys are crucial tools for analyzing patterns in interhousehold transfers of goods and cash, showing how private transfers are related to a household's access to public transfers. Policymakers need to know if government programs are likely to induce changes in private transfers. For example, increasing publicly funded pension benefits may not benefit the elderly as much as expected if their children react to this increase by reducing the amount of private transfers they give to their parents. Survey data can be used to simulate what would happen under different scenarios. Researchers have used household data sets from many developing countries—Colombia, Côte d'Ivoire, Ghana, Kyrgyzstan, Peru, the Philippines, Poland, Russia, and South Africa to study the role of transfers.

In recent research, analysts have found that private transfers are directed toward those households that are often the focus of government benefits—households that include members who are poor, elderly, infirm, unemployed, or without access to formal credit (such as women and young people). Moreover, the research confirms that government policy affects private transfers. Evidence shows that public transfers can "crowd out" private ones. Cox and Jimenez (1993) estimate that, in Peru, an increase of 100 intis in public pension payouts would be associated with a decline of 17 intis in private transfers, leaving a net gain of 83 intis for the elderly household. This phenomenon is most striking in the Philippines, a country with a minimal welfare state and widespread private transfers. A 100 peso increase in public pensions to a retired household was estimated to reduce private transfers by 37 pesos. If unemployment insurance were introduced, private transfers would decline so much that jobless households would be only slightly better off. Although the transfers would still benefit targeted households, the net benefits would be considerably smaller than the gross benefit calculated without reference to the effects on private transfers (Cox and Jimenez 1993).

Determinants of Household Decisions

If the government hopes to influence certain outcomes, such as the number of children enrolled in school, the nutritional status of children, or the number of

children a woman bears, it must understand the factors that influence household decisions. A great deal of analysis of demographic issues has been done using survey data (see, for example, Ainsworth 1989, 1992; Benefo and Schultz 1994; Montgomery and Kouamé 1995; Oliver 1995a, 1995b; and Schafgans 1991). Questions related to fertility that can be investigated with survey data include the following.

- What effects do female schooling, male schooling, and household income have on fertility?
- What factors induce couples to have fewer children and to invest more in each child?
- How do the availability, quality, and price of family planning services affect contraceptive use? What are the socioeconomic characteristics of users and nonusers who have access to public family planning services and those who do not?
- What economic factors affect child mortality? How does child mortality affect the family's fertility decisions?

Using such data to collect information about fertility and contraceptive use in Côte d'Ivoire, analysts found that women in the highest consumption quintile had the lowest age-specific fertility rates, but those in the lowest consumption quintile had the next lowest current fertility. At the same time, current fertility was sharply lower among all women with secondary schooling and among women over age thirty with primary schooling. These data suggest that increasing incomes among the poorest Ivorian women will increase fertility unless levels of female schooling are also raised (Montgomery and Kouamé 1995)

The Benefits and Costs of Using Household Survey Data for Policy Analysis

By now, the reader may be convinced that analyzing household survey data can offer beneficial guidance in making some policy decisions and will be eager to know how much it will cost to reap the benefits of these data. Because a benefit cost ratio for survey-based policy analysis is very difficult to calculate, there is little firm evidence.[3] But consider two contrasting anecdotes in which the benefit cost ratio is relatively easy to guess.

Take, for example, the study of consumer-price subsidies of food products in Tunisia that was presented in the previous section. The analysis in that study was based on existing data on government budgets and on a small household survey conducted in 1993 specifically for this purpose. The costs incurred are fairly easy to quantify—about $55,000 ($33,000 for data collection in the small

special-purpose survey[4] and $22,000 for the analysis of the data and the dissemination of the results to policymakers). On the benefit side, the study identified reforms in the targeting of the food subsidies that, for a given caloric transfer to the poor, reduced the cost of the subsidies by 23 percent, or $74.5 million. Thus, the benefit-cost ratio was on the order of 1,300 to 1. Even if the cost estimate were doubled (to allow for things like unrecorded overtime, the cost of the time of policymakers involved in discussing the study, and items contributed in-kind and off-budget) and the policy analysis were ascribed only a weight of 10 percent in the decisionmaking process with the benefits reduced accordingly, the benefit-cost ratio would be 67 to 1.

Another situation in which the benefit-cost ratio is relatively easy to calculate is one in which a survey (or other data collection effort) has yielded data that have not been analyzed or results that have not been disseminated or results that have been ignored. In these cases, no actual calculations are needed to infer that the benefits are close to zero and that the benefit-cost ratio is discouraging. The reader can probably supply anecdotes of this sort from personal experience.

It is difficult to predict how the mix of high return cases and lost opportunities balance out. A final example helps to illuminate this issue. The government of Jamaica was keen to replicate its good use of policy analysis of the food-stamp program, but analytical capacity in Jamaica was limited. Therefore, the government applied to donors to support a project to improve capacity in social policy analysis. The project was designed to be implemented over five years and was expected to cost $3.4 million ($700,000 a year). Jamaica's annual social sector budget is about $400 million. The capacity-building project would only have to result in reforms leading to gains of about 0.2 percent of the social sector budget to yield a positive return.

This discussion has aimed at helping policymakers identify situations in which sound analysis of household survey data may aid them in making policy decisions. It has shown how analysis can feed into decisions in a range of different ways, indicated the rich and varied policy analyses that can be done with data from national multitopic surveys, and outlined the general virtues and limitations of survey data. If this article helps planners to produce one or two well-done studies with policy impact or to avert a couple of badly designed studies, then its benefit-cost analysis will clearly be positive.

Notes

Margaret E. Grosh is a senior economist in the Development Research Group of the World Bank. Many people contributed to this article. Judy Baker, Lionel Demery, Polly Jones, and Kathy Lindert contributed to the understanding of policymaking in Ghana, Peru, Guyana, and Tunisia, respectively. Martha Ainsworth, Paul Glewwe, and Emmanuel Jimenez drafted

synopses of their research for the section on topics suitable for analysis. Angus Deaton, Paul Glewwe, Emmanuel Jimenez, Diane Steele, Jacques van der Gaag, and anonymous referees provided useful comments on an early draft of the article.

1. Sometimes the goal of the achievement testing is not to study how schools induce students to learn but to study how learning affects labor force participation, job choice, or productivity. In this case, adult achievement measures are required, and testing in households rather than schools is more pertinent.

2. Although in general, consumer food price subsidies may affect households' decisions about whether and how much to farm, that is not pertinent in the Tunisian case. The subsidies studied are explicit and financed out of general government revenues. Producer prices are not reduced below border parity prices.

3. To calculate the benefits, one must first be able to calculate the monetary value of a gain in efficiency or equity due to a specific policy change. Occasionally, this is feasible. For example, one may be able to say that after a change in the targeting mechanism of a program, the minimum budget required to deliver *xx* dollars of services to the target group declined by *yy* dollars. Quantifying benefits is usually harder, however. A given reform may have multiple objectives that act in opposite directions. Moreover, reforms often include goals that are difficult to measure in dollar terms, such as increases in transparency, community participation, or sustainability. Second, one must be able to determine what weight the quantitative policy analysis played in the decision to reform the program. If it were wholly responsible, all the benefits would be counted as a benefit in the benefit-cost calculation, but if the analysis only played a minor role in the decision, then its share of the benefits should be discounted. One must also be able to determine the cost of the policy modeling. The costs of the analysts' time will be relatively easy to quantify, as will the costs of any special data collection effort mounted for the exercise. Most good policy analysis, however, relies at least in part on data collection efforts that go on irrespective of the particular policy modeling being evaluated (for example, administrative records, budgets, and ongoing surveys). What share of these costs should be included?

4. Note that this is less than the costs of many new, multipurpose household surveys. The median cost for ISMS projects is about US$750,000, although the range is from less than US$100,000 to US$3.1 million. The large variation depends on factors such as how much capacity building is incorporated into the project, how much technical assistance is used, how many vehicles are purchased, the size of the sample, the length of the questionnaire, and local prices. For more on costs for ISMS surveys, see chapter 8 of Grosh and Munoz (1996). Costs for SDA Integrated Surveys are of the same order of magnitude (see Delaine and others 1992).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Ainsworth, Martha. 1989. *Socioeconomic Determinants of Fertility in Côte d'Ivoire.* ISMS Working Paper 53. Washington, D.C.: World Bank.

———. 1992. *Economic Aspects of Child Fostering in Côte d'Ivoire.* ISMS Working Paper 92. Washington, D.C.: World Bank.

Anderson, Patricia. 1989. "Levels of Poverty and Household Food Consumption in Jamaica in 1989." Institute of Social and Economic Research, University of the West Indies, Mona, Jamaica. Processed.

———. 1993. "The Incorporation of Mothers and Children into the Jamaican Food Stamp Programme." Working Paper 6. Jamaican Poverty Line Project and Planning Institute of Jamaica, Kingston. February. Processed.

Benefo, Kofi, and T. Paul Schultz. 1994. ***Determinants of Fertility and Child Mortality in Côte d'Ivoire and Ghana.*** LSMS Working Paper 103. Washington, D.C.: World Bank.

Blank, Lorraine, and Margaret Grosh. 1996. "Building Policy Analysis Capability Through Living Standards Measurement Study Surveys." World Bank, Policy Research Department, Washington, D.C. Processed.

Cox, Donald, and Emmanuel Jimenez. 1993. "Private and Public Safety Nets: Transfers between Households." *Outreach* 13(September). World Bank, Policy Research Department, Washington, D.C. Processed.

Delaine, Ghislaine, Lionel Demery, and others. 1992. *The Social Dimensions of Adjustment Integrated Survey. A Survey to Measure Poverty and Understand the Effects of Policy Change on Households.* Social Dimensions of Adjustment Working Paper 14. Washington, D.C.: World Bank.

Deaton, Angus, and Dwayne Benjamin. 1993. "Household Welfare and the Pricing of Cocoa and Coffee in Côte d'Ivoire: Lessons from the Living Standards Surveys." *The World Bank Economic Review* 7(September):293–318.

Gertler, Paul, and Jacques van der Gaag. 1990. *The Willingness to Pay for Medical Care: Evidence from Two Developing Countries.* Baltimore: The Johns Hopkins University Press.

Glewwe, Paul, and Gillette Hall. 1994. "Poverty, Inequality, and Living Standards during Unorthodox Adjustment: The Case of Peru, 1985–90." *Economic Development and Cultural Change* 42:689–717.

Gordon, Derek. 1989. "Identifying the Poor: Developing a Poverty Line for Jamaica." Working Paper 3. Jamaican Poverty Line Project and Planning Institute of Jamaica, Kingston. Processed.

Grosh, Margaret E. 1991. *The Household Survey as a Tool for Policy Change: Lessons from the Jamaican Survey of Living Conditions.* LSMS Working Paper 80. Washington, D.C.: World Bank.

———. 1995. "Five Criteria for Choosing among Poverty Programs." In Nora Lustig, ed., *Coping with Austerity: Poverty and Inequality in Latin America.* Washington, D.C.: Brookings Institution.

Grosh, Margaret E., Kristin Fox, and Maria Jackson. 1991. *An Observation on the Bias in [illegible] Based Estimates of Malnutrition Rates.* [illegible] Working Paper 649. World Bank, Latin America and the Caribbean Regional Office, Technical Department, Washington, D.C.

Grosh, Margaret E., and Paul Glewwe. 1995. *A Guide to Living Standards Measurement Study Surveys and Their Data Sets.* LSMS Working Paper 120. Washington, D.C.: World Bank.

Grosh, Margaret E., and Juan Munoz. 1996. *A Manual for Planning and Implementing the [illegible] Survey.* LSMS Working Paper 126. Washington, D.C.: World Bank.

[illegible] Larry. Forthcoming. "Effective Dissemination of Household Surveys." World Bank, Development Research Group, Washington, D.C. Processed.

Montgomery, Mark, and Aka Kouame. 1995. "Fertility and Child Schooling in Côte d'Ivoire: Is There a Tradeoff?" In Mark Montgomery, Aka Kouame, and Raylynn Oliver, eds., *The Tradeoff between Numbers of Children and Child Schooling: Evidence from Côte d'Ivoire and Ghana.* LSMS Working Paper 112. Washington, D.C.: World Bank.

Moser, Caroline, and Jeremy Holland. 1995. "Urban Poverty and Violence in Jamaica: Summary of Preliminary Findings." Paper presented at World Bank Stakeholder Workshop, Kingston, Jamaica, sponsored by the Urban Development Division, Washington, D.C. Processed.

Newman, John, Steen Jorgensen, and Menno Pradhan. 1992. "How Did Workers Benefit?" In Steen Jorgensen, Margaret Grosh, and Mark Schacter, eds., *Bolivia's Answer to Poverty,*

Economic Crisis, and Adjustment: The Emergency Social Fund. Washington, D.C.: World Bank.

Oliver, Raylynn. 1995a. *Contraceptive Use in Ghana: The Role of Service Availability, Quality, and Price.* LSMS Working Paper 111. Washington, D.C.: World Bank.

———. 1995b. "Fertility and Child Schooling in Ghana: Evidence of a Quality/Quantity Tradeoff." In Mark Montgomery, Aka Kouamé, and Raylynn Oliver, eds., *The Tradeoff between Numbers of Children and Child Schooling: Evidence from Côte d'Ivoire and Ghana.* LSMS Working Paper 112. Washington, D.C.: World Bank.

Schafgans, Marcia. 1991. *Fertility Determinants in Peru: A Quantity-Quality Analysis.* World Bank Discussion Paper 116. Washington, D.C.

Simonpietri, Antoine, and Robert Ngong. 1995. "Poverty Monitoring and Analysis in Sub-Saharan Africa: Information Systems Status Report." World Bank, Africa Technical Department, Washington, D.C. Processed.

STATIN (Statistical Institute of Jamaica) and World Bank. 1988. "Preliminary Report: 1988 Living Conditions Survey, Jamaica." Kingston. Processed.

STATIN and PIOJ (Planning Institute of Jamaica). 1989. *Preliminary Report: 1989 Living Conditions Survey, Jamaica.* Kingston.

Synergie. 1996. "How to Promote Better Use of Data in the Framework of Poverty Reduction Programs: Synthesis Report." Paper drafted for World Bank, Institutional and Social Policy Group, Africa Region, Washington, D.C. Processed.

Tuck, Laura, and Kathy Lindert. 1996. *From Universal Food Subsidies to a Self-Targeted Program: A Case Study in Tunisian Reform.* World Bank Discussion Paper 351. Washington, D.C.

World Bank. 1993. "Indonesia: Public Expenditures, Prices and the Poor." Report 11293-IND. Indonesia Resident Mission, Country Department III, East Asia and Pacific Region, Washington, D.C. Processed.

———. 1994. "Guyana: Strategies for Reducing Poverty." Report 12861-GUA. Latin America and the Caribbean Region, Country Department III, Washington, D.C. Processed.

———. 1995a. "Ecuador Poverty Report, Part I: Components of a Poverty Alleviation Strategy." Report No. 14533-EC. Country Department III, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1995b. "Jamaica Survey of Living Conditions (JSLC) 1988–92: Basic Information." Policy Research Department, Washington, D.C. Processed.

———. 1995c. "Viet Nam: Poverty Assessment and Strategy." Country Department I, East Asia and Pacific Region, Washington, D.C. Processed.

Formal Water Markets: Why, When, and How to Introduce Tradable Water Rights

Mateen Thobani

n most countries the state owns the water resources and the hydraulic infrastruc-ure, and public officials decide who gets the water, how it is to be used, and how uch will be charged for it. But costly inefficiencies in the supply and use of water upport a shift from government provision to a market-based approach that is more ffective and less wasteful.

Markets can allow rapid changes in allocation in response to changing demands or water and can stimulate investment and employment as investors are assured of ccess to secure supplies of water. Because of water's unique characteristics, such markets do not work everywhere, nor do they resolve all water-related issues. By esigning appropriate water laws and regulations and by strengthening private and ublic institutions to administer them, formal water markets can effectively address wing demands for groundwater and for water found in rivers, lakes, and canals. essons from Chile's experience demonstrate that formal water markets can improve he economic efficiency of water use and stimulate investment.

n many developing countries, governments consider water to be too precious a esource or too difficult a commodity to be left to the market. Decisions about who gets water, at what price, and for what use have thus been entrusted to public officials. Although the state retains ownership of this asset, it typically ndows private and public entities, such as farmers, industrial users, and power and water companies, with the right to use surface water or groundwater for a articular purpose. These rights are defined in a variety of ways and have been written into law or have evolved through custom (Sampath 1992). The alloca-ion of water rights is typically the responsibility of the government, as is the onstruction, ownership, and operation of the infrastructure such as dams, res-ervoirs, and canals. In some countries that government even installs and oper-

World Bank Research Observer, vol. 12, no. 2 (August 1997), pp. 161–79

ates wells. Recently many countries have transferred operations and management responsibility to user associations; enforcement is the responsibility of public authorities or water user associations—or both. User associations are typically informal entities that play a role in distributing the water of a canal or river. In some countries, however, such as Chile, Mexico, and Peru, user associations are legally recognized bodies that set and collect fees for operating and maintaining the hydraulic infrastructure. Water companies and industries may belong to user associations, but most associations are made up primarily of farmers.

Publicly administered systems of water rights have all too often resulted in inefficiencies in the use and supply of water. Despite its growing scarcity and increasingly costly hydraulic infrastructure, water is often used wastefully. This is especially true in agriculture, which accounts for 70 percent of water use worldwide, compared with only 8 percent used for household consumption. It is not unusual in arid areas to find cities rationing water even as neighboring farmers grow low-value, water-intensive crops using inefficient irrigation technologies. Although governments normally reserve the right to reallocate water toward more desirable uses, in practice they have found it difficult to do so. Even in the face of rapidly changing demands for water, attempts to shift supplies from agricultural users to domestic urban consumers have often generated conflicts and fomented social disruption.

Moreover, government control has not been effective at ensuring that the poor have access to water. In many cities in developing countries, the poorest are not served by piped municipal water and must resort to buying water from private vendors at prices that are several multiples of those paid by better-off residents. And farmers who are politically influential manage to get easier access to water rights, which are obtained without charge and for whose use farmers typically pay only a nominal fee. Because farmers are unable to sell any surplus water, there is little incentive to conserve water by better soil or water management or by growing less water-intensive crops.

Nor has government control over water been effective at maintaining water or soil quality or protecting ecosystems in most developing countries. The discharge of municipal and industrial wastes, the runoff of agricultural chemicals and poor land-use practices in agriculture, mining, and forestry have led to widespread degradation of land and water resources. Water-borne diseases cause an estimated three million deaths annually and render sick a billion more. In addition, poorly designed public irrigation projects and insufficient incentives for water conservation have resulted in extensive soil salinization (a process in which salts impregnate soils, making it unsuitable for agriculture) and contributed to ecological disasters in many countries (World Bank 1993). In many countries water from underground aquifers is pumped out at unsustainable

levels, threatening the livelihood of many agricultural and nonagricultural users, as well as seriously damaging ecosystems.

A system of publicly allocated water rights makes the private sector reluctant to invest in hydraulic infrastructure—or in activities whose operation requires large quantities of water. If the water in a reservoir built with private funds can be commandeered by the government to meet social or political objectives, investors will have little incentive to provide such infrastructure. If water can be reallocated from agricultural to urban users when shortages occur, investors are unlikely to invest in agriculture.

At the same time, public investment in hydraulic infrastructure has often been a losing proposition. There is no dearth of taxpayer-financed dams that were ill-conceived and, because of budgetary difficulties, long delays, and cost-overruns, cost far more than their eventual benefits were worth. Similarly, efforts to construct and operate public wells have also been unsuccessful. In many countries sizable shares of the public infrastructure budget have gone toward expensive hydraulic projects with low or negative economic rates of return (Holden and Thobani 1995). Despite high construction costs, many of these projects have not been adequately operated or maintained, leaving users, the bulk of whom are farmers, without a reliable supply of water and unwilling to pay higher water charges (Ostrom 1992).

The future looks bleak. Demand for water is rising, and because countries have already exploited the less expensive sources of supply, the cost of developing new supplies has increased sharply. Furthermore, as governments face increasing fiscal pressures, they are no longer willing to spend vast public resources to build new dams and wells or even to operate and maintain existing hydraulic infrastructure.

Market-Based Instruments for Water Allocation

Recent approaches to meet the growing demand for water have focused on incentive mechanisms to conserve water and reallocate it among competing uses rather than on developing new sources of supply. At the same time, governments are seeking alternative ways to finance the operation and management of existing infrastructure and the construction of new infrastructure. These alternative approaches may be divided into three categories: pricing policies, informal water markets, and formal water markets.

Pricing Policies

By raising the user price of water to reflect its true scarcity, or opportunity cost (that is, the price the marginal user is willing to pay), authorities hope to induce

users to conserve water, making it possible to divert supplies to higher value uses (World Bank 1993).[1] In principle, if irrigation water near a city could be priced at what a water company would be willing to pay for the crude water (adjusting for conveyance costs), some farmers would give up farming, and others would switch to more efficient irrigation or grow less water-intensive crops. The higher charges would free up water that could be transferred to the water company for treatment and subsequent sale. They would also generate fiscal resources that could be used to improve the performance and maintenance of the existing infrastructure or to invest in new infrastructure.

Serious practical and political problems, however, have prevented any government from pricing water at its opportunity cost. Even if governments could find an inexpensive way to measure and monitor water flow, measuring the opportunity cost of water is difficult because it varies according to location reliability, season, use, and water quality. Reliable and high quality groundwater close to a city that is rationing water will have a high opportunity cost (even after adjusting for pumping charges), while more polluted water along a more distant river with a variable flow will have a low opportunity cost. Even this difference will vary according to the year and season. In a year with average precipitation, the opportunity cost of water will be far lower than it is in a drought year, just as it will be lower in the rainy season than it is in the dry season. Water used for agriculture or mining will have a much higher opportunity cost than water that is used for hydropower generation and is returned to the river in roughly the same quantity and quality. But if water for hydropower is stored during periods when it is needed for other activities such as agriculture, its opportunity cost could be significant.

The political problems are even more intractable. It is politically difficult to charge a farmer for water from a river that serves a town (and therefore has a high opportunity cost) a higher price than a farmer using water from a river that is not near a town. Similarly, it is difficult to charge profitable hydropower companies less than poor farmers. Strong farmer lobbies typically pressure politicians to keep water charges well below their opportunity cost.

Another problem in pricing irrigation water at its opportunity cost is that the price of land already embodies the price of water rights. In areas of low rainfall irrigated land may sell for ten times the price of unirrigated land, reflecting the expectation that the owner of irrigated land will receive water at a low charge. If charges are later raised to reflect the opportunity cost of water, this land will be valued the same as unirrigated land, resulting in an effective expropriation of the farmer's assets. Although government actions frequently alter the value of private assets, the sheer magnitude of asset expropriation implied, the numbers of people affected, and the socially disruptive aspects (in agricultural unem-

ployment) of such a policy make it highly unlikely that opportunity cost pricing can be introduced within a reasonable time frame.

A unique problem affecting water pricing involves "return flows." When a farmer waters crops, only part of the water is absorbed by the plant. Depending on the efficiency of irrigation, a significant share of the water—the return flow—will seep underground. This water may enter an underground aquifer and be pumped up by another user, or it may even rejoin the river and be diverted into a canal. If water were priced volumetrically, according to what was received rather than what was actually consumed, farmers using inefficient irrigation (thereby inadvertently helping out downstream users) would pay too high a price. These pricing difficulties do not mean that water should be provided at no charge, but they do suggest that setting water prices administratively to ensure rational use will be difficult in practice and that the consequences could well be socially and politically disruptive.

Informal Water Markets

Where governments have failed to respond to rapidly changing demands for water, local (spot) water markets have emerged in several water-scarce countries. Although such informal markets, in which users contract for water on their own, are technically illegal, governments usually turn a blind eye to them, perhaps because they manage to reallocate water quickly and voluntarily. In a typical transaction, a farmer sells a specified volume of his surplus groundwater or surface water for a season or a specified period to a neighboring farmer. Or several farmers collectively sell some of their water to a nearby town.[2]

In this way water is reallocated to more valuable uses without penalizing existing holders of water rights. At the same time the ability to sell provides an incentive for conserving water and using it more rationally. Such informal markets are widespread in South Asia (Pakistan Water and Power Development Authority 1990; Shah 1991, Saleth 1996). The author found them to have been widespread in Mexico even before the introduction of formal tradable water rights. (See also Meinzen-Dick 1996, who found that informal water markets in South Asia were able to increase poor farmers' access to water).

In some cases these trades have not performed well and have resulted in an economically inefficient allocation of water. In parts of South Asia, wealthier farmers with deep wells charge neighboring smaller farmers a high "monopoly" price for water. As a result crop output is lower than it would be if the water were priced at its opportunity cost—and income inequality is exacerbated. The opportunity to sell such a valuable resource also increases exploitation of groundwater, which can deplete underground aquifers (Saleth 1996). Moreover, be-

cause such transactions are illegal, it is difficult to enforce adherence to the entitlement (water right) and protect the aquifers.

A further complication is that these illegal markets may allow upstream users to sell more than they actually consume (because they may sell the return flow component of their water right), thereby infringing upon the rights of third parties. In addition the buyer lacks the security of an enforceable contract. Trades are therefore limited to spot sales or to sales for a single season, often between neighbors; longer-term trades are nonexistent, depriving potential investors or water companies secure long-term access to water. Finally, compared with opportunity-cost pricing, informal markets do not generate fiscal revenues, nor do they provide sufficient incentives or means for the creation of new infrastructure.

Formal Water Markets

Several governments have established legal tradable water rights in an attempt to retain and extend the advantages of informal water markets while reducing some of the negative costs stemming from their illegal status. The potential to sell water rights makes them more valuable and provides an incentive for conserving water and reallocating it to higher-value uses. In this sense, the outcome is similar to that under opportunity-cost pricing. Tradable water rights also allow leasing of water (for a season, say) and spot sales; in fact, they facilitate such transactions. Finally, by allocating initial water rights, without charge, to existing users or holders of water rights, tradable water rights can circumvent the political problems associated with raising water prices and setting nonuniform charges. Governments can monitor operations and more effectively enforce laws and regulations aimed at preventing the abuse of monopoly power at ensuring that sales do not negatively affect the water available to third parties (that is, at addressing the problems of return flow), and at protecting the environment.

Chile and Mexico are the only countries that have established formal regimes of tradable water rights at the national level, but many of the western states of the United States and some states in Australia have such systems (see Rosegrant and Gazmuri 1995 and Pigram and others 1993 for descriptions of the California and Australian systems, respectively.) There are also pockets of semiformal water markets in some countries, where well-regulated water markets have existed for more than a hundred years, even though they are not consistent with national and state water laws (see Kemper 1996). In Australia and the United States, concerns about the environment and protecting third parties have led to many restrictions on water trading that have added to the cost of transactions—or blocked potentially beneficial trades. For example, water

markets in some areas of the United States, such as Colorado and New Mexico, which have few restrictions on trading, have functioned quite well for more than a century, but restrictions on trading have limited the usefulness of the water markets in California. As a result, farmers in California continue to grow low-value, water-intensive crops even when neighboring cities face water shortages and rationing. The Chilean and Mexican water market regimes are probably better models for developing countries facing water shortages.

CHILE. Under Chile's 1981 water code, the state grants existing water users property rights to both surface water and groundwater without charge.[3] These rights are separate from the land and, except for a few restrictions, owners may sell them to anyone for any purpose at negotiated prices. These water rights may also be leased, used as collateral, and inherited. Interested parties may petition to obtain new and unallocated water rights. If others are interested in the same rights, they will be sold at auction; if not, the petitioner will receive the rights without charge.

Rights are obtained by being recorded in a public registry as either consumptive or nonconsumptive, permanent or temporary (contingent). Nonconsumptive rights oblige the holder to return the same volume of water to a specific location (this right is useful mainly for hydropower generation). Temporary rights can be exercised only if all permanent rights have been met (useful when storage capacity exists). The rights are defined volumetrically (either in liters per second or in cubic meters), but revert to proportional rights (share of the streamflow or of the volume in a reservoir) if the available water does not permit all volumetric rights to be honored. This last feature is particularly important in Chile because few rivers have dams or reservoirs for storage.

Water users associations are responsible for monitoring, distributing, and enforcing water rights at the level of the river basin, primary canal, and secondary or tertiary canal. They own and operate the bulk of the hydraulic infrastructure and set water tariffs. The government continues to manage the headworks for some large dams and reservoirs and to charge users for these services.

MEXICO. Under Mexico's 1992 water law, users may convert their existing nontradable water rights to more secure long-term concessions (with a typical maturity of thirty years). These concessions may be leased or sold as long as the sale does not negatively affect the water rights of other users. For nonagricultural users, farmer associations, and groundwater users, the rights are recorded in a public registry and are defined volumetrically. Because any deficits or surpluses are allocated proportionately, however, the rights are effectively proportional. For individual farmers using surface water, the rights are defined only in terms of area to be irrigated and are registered by the water users association.

Users may forfeit their rights if the water is not used efficiently or if it has not been used for three years. Thus, in principle, these rights are less secure than those in Chile. But Mexico's National Water Commission does not specify what constitutes inefficient water use, and the author found no evidence of users having forfeited their rights for this reason. Although the rights cover a fixed period, the water commission plans to renew the concessions for only a nominal administrative charge. Therefore, in effect, the rights are longer-term and more secure than they appear.

Why Establish Tradable Water Rights?

Tradable rights offer water owners an incentive to sell or lease part or all of their holdings to those who have higher-value uses for it. Tradable rights give buyers of water a strong incentive to conserve water to keep their costs to a minimum. Farmers who are considering purchasing new rights are thus more likely to use efficient irrigation techniques, and water companies are more likely to try to reduce water losses.

Such voluntary and flexible transfers that divert water to more productive uses have indeed occurred in the water-scarce areas of Mexico and Chile. In Mexico the transfers benefited some small farmers whose unprofitable farming activities had led to the accumulation of unsustainable debt. In the past they would have resorted either to selling their land and water rights, which would have forced them to leave the land, or to illegally selling their water rights at a lower price. Because they can sell the rights legally, however, and thereby obtain a better price, they have been able to use the proceeds to pay off their debts. Some have even been employed by the farmers that bought the water rights. Similarly, firms that formerly resorted to extracting groundwater illegally have begun buying groundwater rights legally, thereby reducing the problem of aquifer depletion while obtaining secure water rights.

In Chile farmers sold or leased their surplus water rights to more efficient neighboring farmers, industrial users, or water companies. The sales and leases have allowed some water companies and industrial users to obtain reliable access to water without expensive infrastructure investment. The results, according to one study (Hearne and Easter 1997), have been large gains to society. For example, the city of La Serena was able to purchase 28 percent of its water rights from neighboring farmers, allowing the government to postpone the construction of a proposed dam. Similarly, the city of Arica, in the arid north, has been able to meet the needs of urban residents by leasing groundwater from farmers. Such measures have contributed to Chile's success in providing water to virtually all urban residents.

Changes in the structure of water markets create new opportunities for conserving water. When Santiago's municipal water company, EMOS, was notified that it could no longer receive new water rights without charge, the company initially sought to purchase additional water rights. When potential sellers demanded too high a price, FMOS decided instead to rehabilitate its aging pipe structure to reduce water leakages. Similarly, farmers who must pay for water rights to expand production have an additional reason to install efficient irrigation, to use better soil management techniques, or to grow less water-intensive crops. Some Chilean farmers have even used options contracts as a way to avoid buying water that they might not need. The option allows a farmer (who needs to ensure that his trees, say, do not die if there is a drought) to pay a neighboring farmer growing an annual crop for the option of buying water at a prenegotiated price in case of a drought.

The conservation occurred even though water charges were not raised. In fact, several water users' associations in Chile reported that water charges fell after the new water law was passed in 1981. Even before the introduction of the current law, the government was charging users the full cost of system operations and maintenance, which was handled by public authorities. Under the new legislation, those who hold water rights are responsible for setting water tariffs as well as for operating and maintaining the infrastructure. Because they have been able to provide these services at a lower cost, water tariffs have declined.

The situation was different in Mexico, however, where water prices were so low that government subsidies for operating the infrastructure amounted to 0.5 percent of gross domestic product. Just before tradable water rights were introduced, Mexico began to move toward charging full recovery of service costs and to turn over operational responsibility to users. Although costs for operations and management fell, the savings were not enough to compensate for the reduction in subsidies, so water users in Mexico are generally paying higher water tariffs.

The Effect on Poverty Reduction

Secure and tradable water rights reduce poverty in several ways. First, they allow scarce resources to be redeployed for more productive purposes, thus leading to increased output and employment. This occurred, for example, when farmers in Chile and Mexico sold their water rights to more productive farmers or cities. Second, tradable water rights encourage new investment in activities that require large quantities of water. An investment in a fruit farm is more likely to be attractive if the investor knows that water will not be transferred to a neighboring city in times or scarcity and that additional water can be pur-

chased from farmers during water shortages. In Mexico investors built a water-bottling plant after negotiating for the water rights from a farmer. Not only was the farmer better off, but the increased investment also generated additional employment. A World Bank (1994) study on Peru found that tradable water rights have the potential to increase private investment in hydraulic infrastructure, freeing up public resources for other activities while allowing rapid and cost-effective development of hydraulic projects.

Third, by empowering user groups to have a say on the issuance or transfer of water rights, secure and tradable rights help protect the poor. When water rights are granted without charge by public authorities, it is typically the rich and politically influential who have easier access to them, often at the expense of the poor. This is the case, for example, in Peru. Fourth, secure and tradable water rights increase the value of the rights, which are often the most precious assets of poor farmers. In Mexico many small farmers were able to take advantage of their ability to sell their water rights while still remaining on the land.

Additionally, by making it easier for cities to obtain water, such markets benefit the poor because they are the most likely urban residents to have been excluded from piped service. Chile provides almost universal coverage of piped water in urban areas. A contributing factor to that extensive coverage is the ability of water companies to obtain "raw" water at a reasonable price (as in Arica and La Serena). In cities such as Lima or Karachi, where municipal water availability is often limited to certain hours of the day, improved availability of raw water would allow households to receive water at any time. Finally, because the transfer of water to higher-value uses occurs without confiscating water from less productive users (farmers) and without having to build new infrastructure, it is cheaper and fairer than alternatives, such as raising water charges substantially.

No one has measured the effect of tradable water rights on economic growth in Chile and Mexico. Anecdotal evidence and studies showing the gains from trading water suggest, however, that water rights have facilitated economic growth. Agriculture in Chile grew 6 percent a year in the decade following the passage of the water law. In Mexico more efficient farmers were able to expand their output substantially by buying surface water rights. Similarly, the purchase of groundwater rights made it possible for industry to expand production and employment. Without the opportunity to sell water rights legally, the adjustments following the peso's drop in December 1994, which led to changes in relative prices and a decline in domestic demand, would have been more difficult.

In principle, inadequately regulated water sales could lead to erosion from deserted land or threaten the environment if minimum flows at the lower sec-

ons of rivers are not maintained. In practice, neither the author nor Hearne nd Easter (1997) found any evidence of such problems in Chile, mainly beause Chilean farmers rarely sold all their water rights, because rivers in Chile re short, with little return flow, and because some water users' associations orbid trades that could reduce the availability of water downstream. In both 'hile and Mexico, trades require the approval of the pertinent users' association as well as the public water authorities. Although Chile's water authorities ecently reported instances of environmental degradation caused by water sales, ich transactions do not appear to have caused serious problems. In fact, by iducing conservation, water markets have postponed the need to build new ifrastructure, such as the proposed dam near La Serena in Chile, thereby averting otential environmental problems.

Vhen to Establish Tradable Water Rights

water markets have all these advantages and have worked well for several ears in Chile, why have most other countries not adopted them?[4] There are any possible reasons, the relative importance of which varies.

- Some countries, for cultural or religious reasons, object to the idea that life-sustaining water should be bought and sold.
- Some fear that rich individuals or companies will buy up all the rights, excluding the poor from access to water and raising equity and monopoly concerns.
- Another concern is that small scale farmers, either in desperation or ignorance, will sell their rights for a pittance and lose their livelihood.
- Some maintain that water transfers will damage the environment by depleting aquifers, increasing water pollution, or changing ecosystems.
- In some cases, the few that stand to gain from the current system may effectively oppose changing it.
- And because water use often has social benefits that exceed private ones, there may be a sense that public control of water is necessary to ensure adequate investment and low prices.
- The final reason relates to the costs stemming from setting up a new legal, regulatory, and institutional framework; from defining, measuring, and enforcing water rights; and from making necessary changes in water intake and in the conveyance infrastructure to effect the transfers. Closely related to these costs are the difficulties of implementing the initial allocation of water rights, of ensuring that sales of water by one user do not affect the water rights of others (the return flow problem), and of establishing or strengthening public and private institutions to permit a well-functioning market. Given

these costs, the potential benefits from trading water must be sufficiently large for governments to consider establishing tradable water rights.

These problems, however, are not unique to tradable water rights; even publicly administered systems of water allocation must cope with them. Water rights are difficult to define, measure, and enforce even when they are not tradable, and institutional arrangements are essential to manage allocation and distribution. Although water markets may require more complex infrastructure, delivery systems to transfer water are needed regardless of the method that is used to reallocate the water. And most publicly administered systems of water rights have not protected the poor. Efforts to keep water charges low have often resulted in poor service and excluded low-income residents even while better-off farmers and wealthier urban residents obtained water at highly subsidized prices. Although water markets can lead to monopolies, experience shows that suitable antitrust and tariff legislation often results in lower prices and a higher level of service compared with government management of the resource. Similarly, where social benefits exceed private benefits, a subsidy may be preferable to government ownership and control.

Formal water markets tend to reduce the extent of the water problems facing countries. For example, by increasing the implicit value of water rights and by empowering users, water markets provide better incentives to define, measure and enforce rights to water, and they strengthen private institutions such as water users' associations. The infrastructure needed to implement the transfers is likely to be less expensive if it is undertaken by the users than by public authorities. Moreover, the users will build the infrastructure only when it is economical to do so. Government provision is often hampered by the demands of vested interest groups that promote costly projects, such as dams, tunnels and other infrastructure, even when the benefits provided by such services are well below the costs of construction.

Before governments consider establishing a costly new legal and institutional framework, certain minimum conditions must be met. First, because of the costs of identifying potential trades and of making, recording, and enforcing changes in water intakes and conveyance infrastructure, water must be quite scarce—and therefore of high value. Additionally, the infrastructure must be flexible enough to allow trades (for example, adjustable gates rather than fixed flow dividers). Second, society must be willing to enact legislation that respects private property rights to water and recognizes such rights for a reasonably long period. Where cultural, constitutional, or religious factors preclude this commitment, effective formal markets are unlikely to develop or to stimulate private investment and improved water use.[5] Third, minimum institutional capacity in the public and private sector must exist or be developed before tradable

water rights can be established. Private institutions such as user associations that operate at the level of a ditch, canal, or river basin are needed to help establish the initial allocation of water rights and to operate the system. Public institutions must establish the legal and regulatory framework to register the rights, to operate parts of the system that users cannot, and to settle disputes that cannot be resolved by user associations. Finally, because of the likely opposition from those with a vested interest in maintaining the status quo, the political leadership must be prepared to withstand the opposition. If these conditions are met for a substantial part of the country, governments should consider legislation establishing tradable water rights.

Introducing Tradable Water Rights

Even when policymakers are convinced of the advantages of tradable water rights, they need to address several issues to ensure a successful outcome. In addition they must be strongly committed and patient. Water allocation is an emotional subject and because of the technical and legal issues involved, the process of changing to a new regime may take several years.[9] Although the design and implementation of tradable water rights need to be tailored to specific country circumstances, the following guidelines, which build upon Chile's successes but avoid its mistakes, may be useful.

Conducting an Information Campaign

Experience shows that it is essential to explain to users and other affected groups the advantages of formal property rights to water. A well-designed information campaign can overcome the opposition to reform by powerful vested interests. The mechanics of trading could be explained by calling on the expertise of representatives from user associations in countries that have successfully introduced formal water markets. Users and other stakeholders can be invited to participate in designing and implementing the legal framework. Discussions and analyses of draft versions of the law demonstrate a willingness to accommodate the concerns of farmers and other users and are essential to successful implementation.

Registering Rights without Charge

The best way to ensure support for the law is to assign rights to users, without charge, based on their historic usage. Although this approach may provide a windfall gain to some farmers, it acknowledges that the land price already reflects access to water at low prices and that the government is unlikely to re-

cover directly the capital costs of investment in infrastructure. Because this procedure also rewards users that are taking more than their fair share of water, there may be merit in trying to rectify some of the most egregious wrongs. If the government were to try to use this opportunity to correct all such mistakes or to confiscate all illegally obtained rights, however, chances are good that the legislation will fail and the injustices will continue.

Once the rules are established, the individual registration process should be coordinated by water users' associations using "block-titling" methods, where the government provides titles simultaneously to all users in a geographic area (Holden and Thobani 1995). The establishment of a registry, whose officials are perceived to be honest and who are given an adequate budget, is a high priority. One way to ensure fiscal autonomy is to impose a small registration charge to cover operational expenses. The public media should be used extensively to enhance compliance with registration and to make sure that farmers are fully aware of the consequences of their failure to register or to sell their rights. Where large quantities of water are allocated for hydropower, the government should ensure that the initial assignment of such rights does not affect the historical supply available for downstream users. This may require specifying the minimum volume of water that will be released each week (box 1)

Auctioning New Rights

New and unallocated water rights should be sold at auction in an open and transparent manner, making active use of the public media. Before any auction the government should establish a minimum reservation price and verify that the water is not being used and is not needed for environmental or recreational purposes. Information on prices and volumes of past auctions and on transactions in the area should be made public. Any costs to enter the auction should be kept as low as possible.

Protecting the Rights of Third Parties

Even if the initial allocation procedure protects existing users, subsequent sales could infringe upon the water rights of third parties—the return flow problem described earlier. In countries where the reuse of such flows is substantial, procedures to protect third parties must be instituted before trades are authorized. One way to do this would be to specify that all water rights have both a consumptive and a nonconsumptive element. While the former can be sold without restriction, supplies for hydropower use can be sold only if the distribution does not deprive other users of water. Thus for most transfers within the same water basin for the same use, owners would be free to sell 100 percent of their

Box 1. ***Chile's Misunderstood Water Problems***

Despite its success in improving water use, Chile still suffers from conflicts between hydropower companies and farmers, from water quality problems, and from anticompetitive behavior in electricity generation.

Farmer-Hydropower Conflicts. The problem developed when the privatized hydropower companies were granted water rights that failed to set out their obligation to release specified volumes from the reservoirs when farmers need them. The problem was exacerbated when the companies requested additional water for nonconsumption uses from rivers where consumption rights had not yet been assigned (Rios and Quiroz 1995). This experience underscores the need to specify weekly releases of nonconsumption rights— whether the rights are tradable or not is irrelevant.

Water Quality. These problems reflect Chile's failure to enact regulations to enforce the high standards set in the environmental law. The security or tradability of the rights is not an issue.

Anticompetitive Behavior. Although not directly related to water markets, a shortcoming in Chile's water code has allowed one power company to obtain nonconsumption rights to most rivers, in a bid to keep out competition. Moreover, the company did not have to pay for these rights because the auction operates under rules that grant the rights to the petitioner without charge if no other parties approach. By establishing a minimum reservation price, by taxing holdings of water rights (analogous to land taxes), and by passing appropriate antitrust legislation, governments can prevent the abuse of monopoly power.

water rights. But if a farmer were to sell his rights to a water company whose return flows do not return to the same aquifer or river, he could sell only that amount which did not return to the aquifer or river.

Because of the technical difficulties in calculating the return flow component on a case-by-case basis, this approach may not be appropriate for developing countries. But it may be possible to calculate averages that specify the volume of water consumed by a certain crop or activity. In those cases in which return flows are an issue, this published volume would become the limit on the amount that owners could sell to buyers (Holden and Thobani 1995). This procedure would work for both surface water and groundwater. Even though this system has shortcomings, it would be a vast improvement over prohibiting all transfers or having no controls, as is the case with informal water markets.

In addition to these hydrologic effects, there may be other important third-party economic effects. It is neither feasible nor desirable to protect against all of these effects, but two points warrant attention. First, when sales of water from one canal system to another result in a loss of water tariff income to a water user association, it may be desirable to compensate the association for some of that lost income. For instance, in the La Lagunera region of Mexico, buyers must pay 70 percent of the water tariff to the original association and 30 percent to the new association. Second, where municipalities lose significant revenues when water rights are trans-

ferred to other regions, arrangements can be made to pay property taxes on these rights to the original municipality; alternatively, an exit lump-sum tax may be worth considering. This is an important issue in some western states of the United States where rural municipalities lose their revenue base when irrigation water is transferred to cities.

Addressing Monopolies through Taxes and Legislation

Because of the large number of owners of water rights and the high prices they will demand to sell those rights, a monopolistic structure in consumptive water rights (agriculture, industry, and so forth) is unlikely—with two exceptions. First, a monopoly could occur when governments auction new water rights, as happened in Chile (box 1). An appropriate minimum reservation price in auctions should help protect against this outcome. This arrangement could be accompanied by taxes on both consumptive and nonconsumptive rights. A tax on the rights, and not on the purpose for which the water is used or the quantity of water used, has desirable characteristics similar to those of land taxes: it does not distort production decisions, and it helps recover public investment costs in infrastructure. The level of the tax could, at a minimum, be set at the difference in land taxes between irrigated and unirrigated land. For reasons of equity and administrative efficiency small holdings of water rights should be exempt from this tax. Second, monopolies could occur when awarding large volumes of new water rights in the process of privatizing public hydraulic projects under construction. To protect against this risk, an appropriate regulatory framework for each hydraulic project being privatized should be developed (World Bank 1994). In addition antitrust legislation could help protect against any possible monopolies arising from the auction of new water rights, such as in the granting of nonconsumptive rights for hydropower generation.

Water Pollution and Aquifer Depletion

Water quality standards or their enforcement need not be changed when establishing tradable water rights. If the standards need revision or enforcement needs to be improved, such measures can be introduced independently. But water markets could exacerbate aquifer depletion because they provide an additional incentive to pump more water from the ground. Formal water markets provide a feasible way to protect against aquifer depletion, because the law would require that groundwater rights be registered. In regions where aquifer depletion is a concern, groundwater users could form an association that would work with government officials to monitor the level of the water table and each

others' use. If the aquifer were not recharging adequately, the user association would decrease the extraction limits of its users proportionately. Moreover, if exploitation by a user resulted in a shortage of water available to others who are legally entitled to it, public authorities could establish proportional reductions in volumetric rights and bar new exploitation. Such a system is preferable to most existing regimes, whereby owners of the land above an aquifer have full rights to its water, even if their use results in its depletion.

Conclusion

Publicly owned water allocation systems have recorded costly inefficiencies in the supply and use of water. Even in water-scarce areas, water is wasted and public hydraulic projects are poorly conceived, implemented, and operated. Moreover, public approaches have failed to protect the environment or to make water accessible to the poor. With increasing populations and budgetary pressures, these water systems are likely to become even more untenable.

Informal water markets, which evolve spontaneously, are politically easy to implement and can lead to improved water use. But, because they are illegal and thus unregulated, they often result in problems. Formal water markets have greater potential for success.

Economic principles and lessons from experience suggest that formal enactment of tradable water rights permits rapid and voluntary changes in water allocation in response to changing demands, thereby improving water use. These formal water markets also increase user participation in allocating water and planning new investments, while allowing businesses to invest in activities that require assured access to water. The resulting increase in employment and income generation can help reduce poverty.

But tradable water rights are not a panacea, and an effective system is not easy to introduce. Chile's experience and the demonstrated superiority of markets over publicly administered means of resource allocation in general suggest that markets are preferable when water is scarce, when the infrastructure to effect transfers exists or can be cheaply developed, when there is a minimum institutional capacity to implement trades, and when there is political will to establish appropriate legislation.

Notes

Mateen Thobani is senior economist in the World Bank's Latin America and Caribbean Region. The author would like to acknowledge the useful comments of Lorena Alcazar, Ariel Dinar, K. William Easter, Karin Kemper, Larry Simpson, and Ashok Subramanian.

1. Note that a policy of pricing water to cover the full cost of building and managing the infrastructure (the long-run marginal cost) is not optimal if the infrastructure is ill-conceived and built at high cost. In the unlikely event that full cost pricing could be enforced, most irrigators, who typically account for the bulk of water use, would be unable to afford water and would be forced to give up farming. Most of the water would therefore go unused

2. It is useful to distinguish between the water charge paid by a user for system operation and maintenance and the price paid to a seller to use a given volume of water or to lease a certain volume of water for a given period. To draw an analogy from the condominium market, the former is the condominium fee, while the latter is the fee paid for renting the condominium from its owner.

3. The registration process is costly, however; it requires posting announcements in major newspapers, so most small farmers do not actually register their rights. Nonetheless, these rights continue to be honored, but the farmers are unable to sell their rights independently of their land.

4. Because of the limited sales of such rights as well as serious water-related problems, some observers question this statement. In fact, although few trades occur in the high rainfall southern regions and in canals that use fixed flow dividers, many beneficial sales and leases of water are made in the water-scarce north. Also, Chile's water problems are largely unrelated to water markets (see box 1).

5. Mexico chose to get around the constitution by calling water rights long-term concessions but treating them as if they were property rights for land

6. In Peru, despite more than three years of debate and technical assistance, passage of a law establishing tradable water rights is still uncertain

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Hearne, Robert R., and K. William Easter 1997. "The Economic and Financial Gains from Water Markets " *Agricultural Economics* 15 187-99

Holden, Paul, and Mateen Thobani. 1995 "Tradable Water Rights A Property Rights Approach to Improving Water Use and Promoting Investment " *Cuadernos de Economía* 97(December):263–89.

Kemper, Karin. 1996. *Water Resources Allocation and Use in the Curu Valley, Ceará, Northeast Brazil.* Linköping, Sweden: Linkoping University

Meinzen-Dick, Ruth 1996. *Groundwater Markets in Pakistan Participation and Productivity* Research Report 105. Washington, D.C : International Food Policy Research Institute

Ostrom, Elinor. 1992. *Crafting Institutions for Self-Governing Irrigation Systems* San Francisco: Institute for Contemporary Studies.

Pakistan Water and Power Development Authority 1990 *Trading of Canal and Tubewell Water for Irrigation Purposes.* Islamabad: P&I Publications.

Pigram, John J., Robert J. Delforce, Michelle L. Coelli, Vol Norris, George Antony, Raymond L. Anderson, and Warren Musgrave. 1993. "Transferable Water Entitlements in Australia." Report to the Land and Water Resources Research and Development Corp, Australia. Center for Water Policy Research, Armidale, NSW. Processed.

Rios, Monica, and Jorge Quiroz. 1995. "The Market for Water Rights in Chile: Major Issues." *Cuadernos de Economía* 97 (December):317–45.

Rosegrant, Mark W., and Renato Gazmuri. 1995. "Reforming Water Allocation Policy through Markets in Tradable Water Rights: Lessons from Chile, Mexico, and California." *Cuadernos de Economía* 97(December):291–315.

Saleth, R. Maria. 1996. *Water Institutions in India: Economics, Law and Policy.* Institute of Economic Growth. New Delhi: Commonwealth Publishers.

Sampath, Rajan K. 1992. "Issues in Irrigation Pricing in Developing Countries." *World Development* 20(July):967–77.

Shah, Tushaar. 1991. "Managing Conjunctive Water Use in Canal Commands: Analysis for Mahi Right Bank Canal, Gujarat." In Ruth Meinzen-Dick and Mark Svendsen, eds., *Future Directions for Indian Irrigation.* Washington, D.C.: International Food Policy Research Institute.

World Bank. 1993. *Water Resources Management.* A World Bank Policy Paper. Washington, D.C.

———. 1994. "Peru: A User Based Approach to Water Management and Irrigation Development." Report 13642 PE. Country Department 3. Washington, D.C. Processed. Also in Spanish, 1995.

An Exchange on Agricultural Extension Programs

An occasional series of articles and comment, presenting differing views.

The Impact of T&V Extension in Africa: The Experience of Kenya and Burkina Faso

Vishva Bindlish • Robert E. Evenson

Agricultural extension programs based on the Training and Visit (T&V) system are operating in some thirty-one Sub-Saharan African countries. Evidence from Kenya and Burkina Faso shows that T&V management enhances the effectiveness of extension and that such programs support agricultural growth and produce high returns on investments. The data indicate that areas served by extension have higher yields and that within these areas the highest yields are achieved by farmers who participate directly in extension activities. As a result, extension helps to close the gap between the yields attainable with existing technologies and those actually realized by farmers. Closing this gap improves agricultural productivity in the short run, but long-term increases in output in Sub-Saharan Africa will depend on the development of improved technologies that are relevant to local conditions.

Slow growth in agricultural production has been a serious problem in Sub-Saharan Africa, challenging domestic and international policymakers. Accelerating that growth will require important contributions from agricultural extension and research. Although these two interact to complement agricultural growth, their respective roles depend on the general level of agricultural development.

Through the Training and Visit (T&V) system of extension, first introduced by the World Bank in the late 1960s (Benor, Harrison, and Baxter 1984), better production methods and new technologies are being introduced to farmers at the field level. T&V aims at closing the gap between the yields attainable using best-practice technologies and the yields farmers actually achieve. This gap is likely to be particularly large in situations such as those in Sub-Saharan

The World Bank Research Observer, vol. 12, no. 2 (August 1997), pp. 183–201

Africa, where farmers have little education, research and extension are unavailable, and markets and infrastructure are relatively undeveloped. Correcting this gap under these conditions can result in a considerable—and permanent—increase in agricultural growth.

But the role of extension becomes more restrained once most of this gap has been exploited; at this point its contribution depends mainly on the productivity of the agricultural research system and the dissemination of the recommended innovations. Extension thus has considerable potential to make a significant contribution to agricultural growth in Sub-Saharan Africa. But if it is to realize this potential and embrace the vast majority of small subsistence-oriented farmers, the public sector must be involved. Currently, private-sector extension, although valuable in advanced countries, is provided in Africa mostly by commodity organizations and by companies that supply inputs (seeds, fertilizer, and so on) it thus tends to cater to the better-off farmers who produce high-value crops with modern inputs. As a result, not only are small farmers who produce traditional crops unlikely to benefit from private extension, but the economic gulf between poor farmers and those who are more successful may widen.

Extension is a high priority in the World Bank's strategy for accelerating agricultural growth in Sub-Saharan Africa (Cleaver 1993). This strategy is aimed not only at propagating improved practices, but also at helping farmers to become better managers and more adept at organizing their operations and conserving natural resources. As farmers' skills improve and demand for yield increasing research and other services rises, extension services provide the blend of basic science and practical experience essential to stimulate agricultural growth.

Extension programs supported by the World Bank now operate in some thirty one Sub-Saharan African countries. These programs employ about 30,000 people and directly or indirectly reach some 100 million farm households (Schwartz 1993). The World Bank's financial commitments for these operations, which are in addition to the expenditures incurred directly by their governments, amount to almost $450 million (Bagchee 1994). The T&V system provides comprehensive agricultural extension services (for crops, livestock, and so on) within a single line of command. The strategy that has been developed has four key elements regular visits by extension workers to designated contact farmers and contact groups carefully selected to achieve a "spread effect" to farmers who are not in direct contact with extension; a cadre of subject matter specialists, who are trained by research scientists and who in turn train field-level extension workers; the regular supervision of extension staff at all levels; and fixed work programs and specific responsibilities. Extension workers gradually disseminate technological packages to farmers, focusing on a few simple messages on each visit. The initial emphasis is on improving crop husbandry (land preparation, timing of different operations, planting densities, fertilizer application, and so on).

The strategic role assigned to extension—and the resources being invested in it—make it important to evaluate how much extension contributes to production and whether the value of the increase in production justifies the investment. A recent review of the literature on the economic impact of extension concludes that extension services do increase agricultural production but finds limited evidence on the returns on these investments (Birkhaeuser, Evenson, and Feder 1991). Of the forty-eight studies reviewed, only eight reported estimates of the returns on investment. The review found no recent work on the profitability of extension in African countries, and only one study, on India (Feder and Slade 1986), evaluated T&V extension.

This article examines the effect of T&V extension on agricultural production in Kenya and Burkina Faso and attempts to measure the value of the increase attributable to it. The article draws on the results of two evaluations—Bindlish and Evenson (1993), and Bindlish, Evenson, and Gbetibouo (1993). We first review the financial allocations and personnel requirements of the extension programs in both countries as well as the characteristics of the farmers in the sample. We then summarize the results, looking at whether farmers ignore or adopt the advice of extension workers, and examine how effective farmer-to-farmer communication is in delivering extension messages. Finally, we review the impact of extension on production in order to analyze extension's contribution to profitability and growth.

Agricultural Extension Services in Kenya and Burkina Faso

In 1982 Kenya introduced a pilot T&V program in two administrative districts. By 1985 the program had expanded to cover thirty districts, or about 90 percent of the total arable land. The remaining areas are being covered under the second phase of the project, which is expected to operate until 1998. In Burkina Faso T&V was introduced across the country's twelve administrative units, known as CRPAs (Centres Regionaux de Promotion Agro-Pastorale) during a four-year period ending in 1989, at which time it was adopted as the national system.

The study sample for Kenya consisted of 676 farm households selected randomly from seven representative districts (Bungoma, Kericho, Kisumu, Machakos, Murang'a, Taita Taveta, and Trans Nzoia). These seven districts account for about a fourth of Kenya's population, a fifth of its arable land, more than a third of its maize production, and about a fifth of the output of most other crop and livestock products. The farms were spread over all thirteen agro-ecological zones found in Kenya. The study sample for Burkina Faso, which included 3,609 farm households, was much larger and covered all twelve CRPAs.

In both Kenya and Burkina Faso, T&V replaced earlier extension syste
that, while productive, were considered to have certain weaknesses (see, f
example, World Bank 1983). These weaknesses related to deficiencies in t
training of field extension agents and poor supervision of staff. Because the
weaknesses are precisely the kind that the T&V management structure seeks
address, the question was whether T&V had a greater impact on producti
than the extension system it replaced.

To test this hypothesis one must compare the data from a period when T8
programs were operating with a period preceding its introduction. Althou
Kenya had such data, Burkina Faso did not.[1] This meant that a before-an
after framework could not be used in the case of Burkina Faso. But T&V w
phased into Burkina Faso during the four-year period from 1986 through 198
Thus the hypothesis tested instead was that T&V's effect on production, es
mated on the basis of data collected in 1990–91, would be greater in tho
CRPAs in which it had been introduced earlier. A second hypothesis was th
farmers who were members of a T&V contact group were more likely to ado
recommended practices, although other farmers were also considered to ha
benefited.

Resource Allocations

In Kenya extension expenditures increased 19 percent, or 75 cents, a year p
farm household, after the introduction of T&V (table 1). The total numbe
field extension agents in the seven Kenyan districts increased from 774 in 198
to 1,071 in 1990–91, reducing the average number of farm households serv
by each agent from 913 to 816. In Burkina Faso extension expenditures act
ally declined after the introduction of T&V, falling by 29 percent from $10.
per farm household in 1985–86 (the year preceding the introduction of T&

Table 1. *Resources Allocated to Extension*

Allocation	*Kenya*[a] *Before T&V*	*After T&V*	*Burkina Faso*[b] *Before T&V*	*After T&V*[c]
Extension expenditures per farm household (1991 dollars)	3.92	4.67	10.2[illegible]	[illegible]
Average number of farm households per field-level extension worker	913	816	679	1,001

a. Based on data for the seven sample districts. The "before T&V" estimates refer to 1982–83; t
"after T&V" is the average for 1983–84 to 1990–91.

b. The "before T&V" refers to 1985–86; the "after T&V" is the average for 1989–1990 to 1990–9

c. 1990–91 alone.

Source: Bindlish and Evenson (1993); Bindlish, Evenson, and Gbetibouo (1993).

to an average of $7.25 a year from 1989–90 to 1990–91, when T&V was the national system. That drop reflects a government decision to rationalize the use of staff resources by reducing the number of agents employed from 1,005 to 839. As a result the average number of households each agent served rose from 679 to 1,001.

Extension Advice

Farmer groups, traditionally an important feature of rural society in Africa, were the main contact points for extension agents in both Kenya and Burkina Faso. Accordingly, a high proportion of Kenyan farmers who reported receiving extension advice after the introduction of T&V were members of such groups. In Burkina Faso 21 percent of the sample farmers belonged to a T&V contact group.

After the introduction of T&V in Kenya, the proportion of farmers who reported receiving extension advice increased from 6 percent to 48 percent (table 2). Eighty-eight percent of the respondents reported that this was the first time they had been reached by extension services. In Burkina Faso 31 percent of the farmers reported being served by the T&V-based extension system. Twenty nine percent had participated the previous year in scheduled meetings between field agents and their contact groups, but only 21 percent reported actual membership in T&V contact groups; thus it appears that an appreciable number of farmers who were not contact-group members also participated in group activities. Virtually all those who received extension advice considered it to be useful.

Under T&V, extension appears to be reaching female farmers, although the small number of women in the Burkina Faso sample makes it difficult to draw inferences from the data for that country. In Kenya roughly the same proportion of farmers from female-headed households (45 percent) and male-headed households (50 percent) received extension advice, as did similar proportions of small- (45 percent), medium- (45 percent) and large-scale farmers (53 percent). In Burkina Faso only 21 percent of the small-scale farmers and 30 percent of the medium-sized farmers, compared to 47 percent of the large-scale farmers, participated in scheduled contact group activities. This low level of participation suggests that small-scale farmers should be explicitly targeted for membership in contact groups.

Influencing Awareness

Table 3 shows how effective T&V programs were in making farmers aware of new production practices. In Burkina Faso more than 80 percent of the farmers

Table 2. *Sample Characteristics and Extension Advice*
(percentage of sample)

	Kenya	*Burkina Faso*
Sample characteristics		
Contact farmers	4	—
T&V contact group members	31	21
Members of female-headed households	36	4
Attended at least primary school	55	33
Small farms[a]	49	31
Medium farms[a]	34	43
Large farms[a]	18	27
Members of households with agriculture as primary source of income	72	-
Extension advice		
Farmers reporting receiving extension advice since T&V introduced	48	31
Members of male-headed households	50[b]	
Members of female-headed households	45[b]	-
Farmers reporting receiving extension advice before introduction of T&V	6	
Members of male-headed households	7[b]	
Members of female-headed households	3[b]	
Extension advice by farm size		
Small	45[b]	21[b]
Medium	53[b]	30[b]
Large	45[b]	47[b]

— Not available.

a. Small farms are defined as two hectares or less, medium farms are between two and five hectares in Burkina Faso, two and eight hectares in Kenya. Farms more than five hectares in Burkina Faso and more than eight hectares in Kenya are considered large.

b. Refers to the percentage of sample farmers in the indicated category.

Source: Bindlish and Evenson (1993), Bindlish, Evenson, and Gbetibouo (1993).

who were members of T&V contact groups and 50 percent of those who were not said they were aware of the extension messages. (The relatively low level of awareness about improved cultivars might reflect a preference for local varieties or the general paucity of improved crop varieties for environments with low and uncertain rainfall.)

In Kenya awareness of extension messages declined with the complexity of the practice. More than 80 percent of the farmers were aware of basic husbandry practices, such as spacing, planting dates, improved cultivars, and basal fertilizer applications. In contrast, less than a third were aware of advice on top-dressing of fertilizer, plant protection chemicals, and stalk borer control, which are more complex practices.

Table 3. ***Farmer Awareness, Testing, and Adoption of Extension Messages for Selected Practices***

(percentage of sample farmers in the indicated category)

	Awareness			Testing			Adoption		
Category	*All farmers*	*T&V contact group members*	*Non-members*	*All farmers*	*T&V contact group members*	*Non-members*	*All farmers*	*T&V contact group members*	*Non-members*
Kenya									
Spacing	76	—	—	—	—	—	71	—	—
Planting dates	79				-	—	76	-	—
Improved cultivars	80					--	74		—
Basal dressing	100	—					63		—
Top dressing	33		—	—	—	—	10	—	—
Chemical use	12	-				--	10	--	—
Stalk borer control	22	--					11	---	-
Burkina Faso									
Soil preparation	88	96	86	21	34	18	60	71	55
Seed treatment	91	97	89	31	43	28	67	83	56
Improved cultivars	58	78	53	11	24	8	21	38	13
Seed drilling	91	98	89	21	39	16	49	67	38
Organic fertilizer	94	98	93	20	28	18	67	69	63
Chemical fertilizer	87	92	86	20	36	16	32	53	21
Pesticide	66	83	62	8	20	5	21	38	11
Animal draft	95	98	94	18	33	14	34	53	23
Motorized draft	66	82	62	1	3	1	1	1	1
Crop rotation	73	86	70	17	27	15	50	57	44
Erosion control	85	94	82	15	24	13	36	43	29
Agro forestry	59	9	54	8	13	7	14	19	9

— Not available.

Note: Kenyan farmers have a long tradition of organizing into groups to market output, obtain credit, [illegible] and the extension system has sought to work with some of these existing groups. Thus [illegible] are not presented separately for group members and other farmers, as in the case of Burkina Faso. Burkina Faso contact groups have been created specifically for extension purposes.

Source: Bindlish and Evenson (1993); Bindlish, Evenson, and Gbetibouo (1993).

Testing and Adoption of Recommended Practices

Farmers have been encouraged under T&V to test extension recommendations on small areas in their fields before adopting them. Such testing was common in Kenya in the years immediately after T&V was introduced, but the practice has decreased over time. In Burkina Faso, however, a significant amount of testing still appears to take place, particularly among members of contact groups (see table 3). Farmers who did not belong to contact groups reported some testing, but no more than 20 percent of such farmers tested most recommended practices.

The level of adoption of extension advice was lower than the level of awareness. For Kenya 63 to 75 percent of the sample farmers adopted extension recommendations covering the four simpler practices (spacing, time of planting, improved cultivars, and basal dressing of fertilizer), but only about 10 percent adopted the three more complex practices (top dressing, plant protection chemicals, and stalk borer control). The differences in adoption rates between male- and female-headed households, and among small, medium, and large farms, were significant only for fertilizer, which suggests that female-headed households and small-scale farmers face relatively greater financial constraints and may not be in a position to purchase inputs.

Overall adoption rates in Burkina Faso tended to be modest, although they were higher for T&V contact group members than for other farmers. Note that although the twelve practices selected for analysis have wide regional applicability, their relative importance varies by agroclimatic area. Thus, adoption rates were higher in the regions best suited to these practices. For example, more than 80 percent of the sample farmers located in the northern Sahel followed the recommended use of organic fertilizer, and about one-half adopted complex anti-erosion methods (Bindlish, Evenson and Gbetibouo 1993, table 13). These two practices were also adopted by farmers in the central plateau and eastern zones, both of which have serious soil erosion problems. In contrast, in the western and southwest zones, where crop production is important, a high proportion of farmers followed crop rotation (89 percent) and soil preparation (99 percent) practices and invested in chemical fertilizer (80 percent) and improved cultivars (62 percent).

Farmer-Identified Constraints

Kenyan farmers cited the cost of inputs as the key constraint in their efforts to increase the use of purchased inputs, suggesting a possible need to expand access to agricultural credit. Although they identified labor shortages as the main reason for ignoring recommendations concerning spacing and timely planting, that constraint is also likely to reflect financial problems that prevent farmers from hiring labor. Farmers in Burkina Faso also identified the cost constraint as the main reason they failed to use purchased inputs, but they cited insufficient knowledge as the critical factor limiting their adoption of a majority of the other practices. Insofar as high proportions of farmers reported being aware of the extension messages for individual practices, this apparent contradiction suggests that farmers lack detailed knowledge about the correct methods of application. To achieve better results, extension may need to work more closely with farmers to familiarize them with all facets of the messages. Notably, only a small proportion of

the sample farmers in both countries cited the unavailability of inputs as an impediment to adoption.

Sources of Information

The extension service cannot reach all farmers directly, and T&V, like all extension systems, assumes that messages will spread to other farmers from those contacted directly (table 4). Even some of the sample farmers in Kenya apparently first acquired their knowledge of extension messages from other farmers; 48 percent of the sample farmers reported receiving advice from extension agents, but only 25 percent indicated that these agents were their initial source of information. In Burkina Faso, as noted earlier, many farmers who are not members of T&V contact groups attend extension meetings. Forty-one percent of the farmers who were not members of such groups identified other farmers as their source of extension messages. The majority of farmers who were members attributed their knowledge of extension to the T&V agent. In both countries, radio broadcasts were a small, but positive, influence.[3]

If farmers feel that they do not have enough information to invest in new practices, extension may have a larger direct role to play in delivering more complex and specialized messages (see Bindlish and Evenson 1993, table 29, on Kenya; Bindlish, Evenson, and Gbetibouo 1993, table 17, for Burkina Faso.) For instance, more sample farmers in Kenya attributed their initial knowledge of simpler practices to other farmers than to field agents, but for the more complex practices, about the same number attributed their initial knowledge to field agents as to other farmers. In Burkina Faso higher proportions of farmers who were not in T&V contact groups identified extension agents rather than

Table 4. *Sources of Information*

(percentage of sample farmers in the indicated category)

Source	Kenya, all farmers	Burkina Faso		
		All farmers	Contact group members	Nonmembers
Extension workers	25	44	74	36
Other farmers	41	36	14	41
Radio	2	10	7	11
Input suppliers	2	n.a.	n.a.	n.a.
Marketing agents	2	n.a.	n.a.	n.a.
Other sources	5	11[a]	5[a]	12[a]
Not stated	23	n.a.	n.a.	n.a.

n.a. Not applicable.

a. Includes input suppliers and marketing agents.

Source: Bindlish and Evenson (1993), Bindlish, Evenson, and Gbetibouo (1993).

other farmers as their primary source of information on practices using modern inputs (chemical fertilizers, pesticides, and improved cultivars).

Determinants of Advice, Testing, Awareness, and Adoption

Efficiency at the farm level has both technical and allocative components. Farmers must use improved production techniques in combination with a more effective allocation of resources. Because extension is primarily concerned with delivering technical messages, its contribution to economic efficiency is largely through its effect on technical efficiency. How effective has T&V extension been in contributing to technical efficiency in Kenya and Burkina Faso? Using econometric techniques to analyze rates of change, we examined whether T&V had led to earlier and greater awareness, testing (in Burkina Faso), and adoption of improved practices than would have occurred otherwise. (Adoption was based on data aggregated to the area covered by the extension agent for Kenya and to the village level for Burkina Faso.) We also looked at the probabilities of individual farmers becoming aware of and responding to these messages. The focus was on the specific practices for Kenya and Burkina Faso discussed earlier.

Participation in extension activities as contact farmers and members of contact groups would reflect a direct involvement with the T&V system and determine, to an important extent, its impact on farmers' efficiency. The decision to participate is endogenous, however, because farmers decide whether they want to be contact farmers or contact group members. Thus, we analyzed the effect of participation using a two-stage procedure. In the first stage, we statistically "predicted" participation in an analysis of interest in its own right. In the second stage, we included predicted participation as an explanatory variable in the analysis of the determinants of awareness, testing, and adoption.

The decision to participate is influenced by the farmers' demand for information as well as by the available supply of information services. The demand factors used in the model to predict participation comprised farmer characteristics, the state of community services and activities (because participation frequently occurs in a group context), and climatic conditions. Variables focusing on farmers' sex, age, and education were used to represent farmer characteristics. Community characteristics were represented by the community means for farm size, education levels, and age; climatic factors by weather variables. Additionally, in Kenya the community characteristics included indicators of the availability of road and transport facilities, and the climatic factors included indicators relating to topographic features and agronomic potential. (All areas

of Kenya have been classified according to whether they have a high, medium, or low agricultural potential.)

Because the supply of extension services is an important determinant of participation, the variable reflecting this supply should measure the services provided relative to the number of farmers served. The Kenya and Burkina Faso studies rely on the most appropriate and widely used variable for this supply index—the ratio of field workers to farmers in a community. The main appeal of this variable is that farmers generally have no control over the number of field agents employed by the government, so one may assume that increased field services contribute to more information and advice. (A more detailed discussion of this variable and possible endogeneity problems follows in the next section.)

The analysis provides useful insights regarding the types of farmers likely to seek direct contact with extension services. For example, in Kenya farmers from households headed by women or by persons with higher levels of schooling have a higher probability of participating as contact farmers or members of a contact group. This finding conforms to expectations: female farmers are more likely to seek extension advice than male farmers because it can substitute for other factors, such as inputs and credit, to which women might have less access; more educated farmers have a greater appreciation for extension advice and expect to benefit from it. The probability was also higher for farmers in areas with only a small number of farm families per field agent and in areas with a low or medium agronomic potential. Again, these findings conform with expectations that relatively high coverage would encourage participation and that farmers would seek extension advice to compensate for agronomic inadequacies. In Burkina Faso contact group members tended to be younger, full-time farmers with more land. Here, too, farmers in areas with fewer farm families per agent were more likely to be in T&V contact groups.

In analyzing those factors that affect a farmer's decisions, we used three variables: predicted participation, an indicator of the year in which T&V was introduced in an area, and the agent-to-farmer ratio. The reason for using the agent-farmer variable again was to see whether extension staffing had an impact independent of participation. In addition, we also incorporated the effects of learning from other farmers in the form of a learning curve. In Burkina Faso we also considered the relationship between testing and adoption.

The data used in the analysis consisted of a time series for 1979–90 for Kenya and 1984–91 for Burkina Faso and were based on the farmers' recollection of the year in which they had first become aware of, tested, or adopted a practice. Thus, the potential for recall error exists, although a chronology of technology diffusion would have been impossible to develop otherwise. For the agent-to-farmer ratio (defined at the "location" level for Kenya, and the zonal level for Burkina Faso), time series had to be constructed. In Kenya we asked extension

staff in 1991 how many years the agents then posted in a location had been there. Because they were unable to provide information on personnel who had been transferred elsewhere, an adjustment was made for these field agents on the basis of the available district-level data. A similar adjustment was made for Burkina Faso using provincial data to derive time series for extension agents at the zonal level.

The results of the rate-of-change analysis and the farm-level probability analysis were generally consistent for both countries, although the results were stronger for Burkina Faso, where T&V had a positive and statistically significant effect on testing for ten of the twelve practices and on adoption for nine practices. Using participation in T&V contact groups or the variable indicating the year in which T&V was introduced in an area gave similar results, suggesting that the impact of T&V is not confined to contact group members but extends to other farmers. Participation in T&V contact groups in Burkina Faso increased the probability that farmers would test and adopt most of the recommended practices. It also confirmed the positive effect of testing on adoption. Farmers with more land and farmers with higher levels of schooling were more likely to learn from other farmers and to test and adopt new technology.

The findings for Kenya were not as clear-cut; evidence regarding the effects of T&V on farmers' awareness of recommended practices tended to be weak, reflecting perhaps the conceptual problems commonly encountered in collecting such data. Because farmers may be unable to define or express their awareness clearly, evidence that the advice has been adopted is a more effective measure of extension's impact. And, in fact, the findings on adoption were more robust. T&V generally had a positive effect on all the recommended practices, and it had a statistically significant effect on most. The response to three of the four basic practices covered in the analysis—spacing, planting dates, and the use of improved cultivars—was particularly strong. As in Burkina Faso, higher levels of schooling led to more and earlier awareness and adoption.

Measuring the Impact on Production

The effect of T&V on farmers' economic efficiency was measured by the change in farm output.[4] An important procedural issue related to the choice of extension variables. Several different specifications were used in earlier studies reviewed by Birkhaeuser, Evenson, and Feder (1991). These included the number of farmer contacts with the extension system, participation in groups that were in contact with extension agents, adoption of a recommended technology, and the ratio of field extension workers to farmers in a region. With the exception of the last, the other variables reflect actions of or choices made by farmers.

Thus, when more productive farmers seek out extension workers for advice more frequently than less productive farmers, one cannot infer that extension has increased their productivity. And, unless an econometric procedure is devised to correct for this endogeneity, they cannot be used as independent variables. (In the previous section, a predicted participation variable, which took endogeneity into account, was created to explain awareness and adoption, and that variable was available for the productivity analysis.)

Most previous studies of this type, however, relied on a measure of extension supply that is determined by the government and is not a farm-specific variable. As noted earlier, the logical choice is a ratio measuring the supply of field services to a target group of farmers. The supplying staff may use different techniques to reach farmers and may provide more services to some farmers than to others, but it is expected that the more staff services available, the more advice and information farmers will receive and that areas with more services will, on average, be more productive.[illegible]

Governments have been known to locate extension personnel in choice agricultural areas in an attempt to demonstrate the success of extension programs, effectively creating a selectivity bias. Such behavior creates a spurious correlation and causes the estimated relationship between extension and productivity to be biased upward (Evenson and Kislev 1975). Our studies attempted to deal with this potential bias in two ways. First, we examined the strategy behind the geographical expansion of T&V. It appeared that in both Kenya and Burkina Faso, the expansion of T&V management at the aggregate level was random once the programs had graduated from the pilot phase; in Burkina Faso the expansion occurred in all regions at roughly the same rate. Nonetheless, decisions concerning the location of extension agents within districts in Kenya and RPAs in Burkina Faso may still have been biased. Second, we applied statistical tests to determine the endogeneity of the agent-to-farmer variable (meaning that its value is not affected by factors outside the model). The tests did not indicate an upward bias, but the possibility is nonetheless real (see the appendix).

There is also a timing dimension between an agent's year of service and farmers' productivity. Not only is extension advice likely to be reinforced as farmers gain experience, but it also takes time to spread the word among farmers. Thus the impact of extension services provided in a given year will extend into the future. This assumption was incorporated into the agent-to-farmer variable in the form of an accumulated stock of field-agent years.

In Kenya the ideal way to look at the effect of extension after the introduction of T&V would have been to compare differences between the data for 1981–82 and 1990. Such an analysis was impossible, however, because 1981–82 data were available only for a limited number of inputs. Instead, we estimated separate models for each period.

The 1990 specification for Kenya was relatively complete. Aggregate farr production was related to inputs, geoclimatic conditions, weather, farmer an community characteristics, and the agent-to-farmer ratio. Two additional vari ables were used for Burkina Faso: the number of years that T&V had bee operating in an area, and the proportion of farmers in a village who were mem bers of a T&V contact group. In both countries, different econometric tech niques consistently gave similar results. In addition, the estimates for th nonextension variables (land, labor, and capital) appeared reasonable and in spired confidence in the measures of the extension variables.

The results for Burkina Faso showed that important determinants of pro duction were not only the agent-to-farmer ratio, but also the variables measur ing the proportion of farmers in a village who were members of a T&V contac group and the number of years T&V had operated in the area. In Kenya th estimates for the agent-to-farmer ratios were considerably higher for 1990 tha for 1981–82. But the impact of extension may have been underestimated fo 1981–82, given the weaknesses in the data for that year. Furthermore, the analysi did not take into account changes in other complementary factors, such as rura infrastructure (feeder roads), access to agricultural credit, and new technolo gies, which might have improved farmers' ability to respond to the extensio messages.

All the same, there is little to suggest that major increases in the supply of suc factors actually occurred in Kenya between 1981–82 and 1990. Indeed, the stud suggests that farmers lacked access to agricultural credit during this period, whic might have reduced the effectiveness of extension. Similarly, few new technolo gies were introduced at this time, because the agricultural research system wa being reorganized. As for investment in feeder roads, the government seems t have been more concerned with maintaining the existing network than with add ing to it. Therefore, although the difference between the 1990 and 1981–8 estimates cannot unequivocally be attributed to T&V, the system seems likely t have increased the productivity of extension. Notably, the results showed tha subject matter specialists with technical education and research experience an extension agents with knowledge of on-farm research were more effective in rais ing farm productivity than those without such credentials.

Economic Analysis

Policymakers are interested in knowing not only whether extension has in creased agricultural output, but also how the benefits compare to the costs o achieving this goal. What can be learned from experience in these two coun tries? To find out, we computed rates of return for T&V extension in both

countries and for extension services before T&V was implemented in Kenya. The computed rates of return are "marginal," that is, they refer to the social return that would be realized on an additional public sector investment, given the investments already made. The statistical methods applied in the study were not suited to estimating average returns, which in any case are not a practical calculation for on-going programs such as extension.

One caveat applies to these results. The high rates of return reported here are a consequence of improved technologies adopted by farmers who had not previously employed such techniques and whose productivity had been low. The impressive increase in output that resulted from these technological innovations (improved crop management practices, seed, fertilizers, and other inputs) is unlikely to be repeated. Further productivity increases will occur as existing practices are modified and improved, but the large initial gains in output are unlikely to continue.

For Kenya the computations showed a marginal rate of return of more than 100 percent for the T&V-based 1990 extension and 28 percent for the pre-T&V 1981–82 extension. For Burkina Faso the computed marginal rate of return was 91 percent.

What do the results for Burkina Faso imply for growth of agricultural productivity? At the level of extension expenditures in the early 1990s (slightly under 1 percent of the value of production), the effect of additional extension agents is less beneficial than participation in T&V contact group activities. (Estimated crop yields for the contact groups were 29 percent higher than average.) For example, as shown in table 3, about 30 percent of the sample farmers participated in T&V contact-group activities in 1990, four years after T&V was first introduced. If the proportion of such farmers increased in the following six years to 75 percent, which would be about the expected peak, that increase would imply a considerable increase in agricultural production at the end of the ten year period.

Conclusion

The evidence from the evaluations supports the hypotheses that extension programs in Sub-Saharan African countries have a positive long-term effect on agricultural growth and that T&V management improves extension performance. Two elements of this management structure appear to be particularly important. One relates to the discipline of regular extension visits to farmers, and the other to the role of the subject matter specialists. In response to the involvement of these personnel, the extension service is under constant pressure to provide timely and competent advice.

The evaluation for Burkina Faso, which shows that extension made two im portant contributions, is especially revealing. The first contribution is that farme in villages with more extension staff time recorded higher crop yields. The sec ond, which augments the first and is more substantial, is that the higher yield are associated with direct participation in contact group activities. This contri bution has the effect of closing the gap between the "best-practice" yields pos sible with the existing technologies and the yields farmers actually obtain. Th improvement can move the agricultural economy to higher levels of producti ity, beyond which progress will largely depend on how quickly new technolo gies can be developed.

This transition has very significant economic value. It can provide Africa countries "breathing room" to stimulate their agricultural research systems an achieve the kind of growth potential attained in Asian countries through th application of new technologies. This transition will be realized only with well managed, disciplined, and effective extension programs. Previous extensio programs have not been able to mine this source of growth, but contemporar T&V programs appear to be doing so.

It is not clear how many Sub-Saharan African countries still can tap th source of growth. Kenya has probably made most of the transition to the nex stage. It has relatively well-developed markets and infrastructure as well as highe levels of farmer education. The high rate of return estimated for its T&V managed extension system is probably attributable, at least in part, to the clos ing of the gap between best-practice yields and actual yields. Markets, infra structure, and farmer education are less well developed in Burkina Faso, whic has a large gap between actual yields and best-practice yields. Its situation i likely to be replicated in much of Africa.

The high rates of return estimated in these two studies cannot be used t justify expanding the agricultural extension systems in Sub-Saharan Africa t intensity levels above those characterizing other, more successful developin countries. The estimates imply that extension programs have sharply decreas ing marginal effects as the programs expand. The statistical reliability of th estimates also declines for calculations far from the mean of the sample. Per haps of more importance is the fact that the high returns estimated for thes programs are based on two important features of extension programs, only on of which is replicable or permanent. Extension effectiveness based on bette management and design of extension systems produces permanent gains if th better management and design elements are maintained. Extension effectiv ness based on the gains from mining the best-practice potential in economie with very low levels of marketing and related institutions cannot be sustaine for long periods of time.

Appendix. Note on Simultaneity

The following procedure describes the test for simultaneity bias arising from governmental decisions concerning the location and size of extension staff. Note that for the economic analysis, conclusions regarding the economic impact of extension were drawn from measurements of the *supply* of extension services to farmers.

To determine the potential endogeneity of the extension supply variable itself, a three-stage least-squares procedure was used. The variable was defined at the location or village level. In Kenya each location includes several sublocations, which represent the area covered by an extension agent. The extension supply equation included two classes of sublocation variables. Current variables, which were included in the aggregate production function, comprised:

- Sublocation indicator of medium and low potential land
- Sublocation indicators for hilly and undulating land.

Lagged (that is, 1981–82) variables were treated as instruments in the first-stage extension supply equation. The argument for these variables is that they are unlikely to be correlated with the error term in the current extension supply equation. (This is a common argument for lagged extension supply.) These variables included:

- Sublocation proportion of farms with national roads within five kilometers in 1981–82
- Sublocation proportion of farms with bus service within five kilometers in 1981–82
- The number of farms in the sublocation in 1981–82
- The size of the sublocation served (square kilometers)
- Mean household education in sublocation in 1981–82
- Mean household size in sublocation in 1981–82
- Sublocation indicator of extension supply in 1981–82
- Sublocation mean area per farm in 1981–82

For Burkina Faso we did not have comparable data from an earlier period. The first-stage extension supply equation for Burkina Faso included three household-level variables (production, area, and education) that were in the second-stage production specification. The instruments included:

- The number of households in the village (the extension staff variable covered several villages)
- Number of years since T&V was introduced
- Dummy variables for political units (the CRPAs).

T&V was phased into Burkina Faso during a four-year period. The T& coverage each year was similar in different political units. The third-stage pro duction functions did include soil and climate class variables, but not the po litical CRPA dummy variables.

In both Kenya and Burkina Faso, the simultaneity correction produced slightl higher coefficients.

Notes

Vishva Bindlish is with the World Bank's Technical Department for Africa; Robert E. Evenso is with the Department of Economics and the Economic Growth Center at Yale Universi The authors wish to thank Kevin Cleaver and the anonymous referees for comments on earli drafts of this paper, as well as Jock Anderson, Gerson Feder, and Dennis Purcell The autho would also like to thank Madhur Gautam for pointing out an error in the estimation of t agent-to-farmer variable. The economic implications of the corrected estimate are not sub stantially different from those originally reported.

1. The data for Kenya were based on a countrywide rural household budget survey unde taken in 1981–82, the year immediately preceding the introduction of the T&V program I 1990–91, when the data for the evaluations were collected, the farmers from the 1981 8 survey were randomly resurveyed. The 1981–82 survey was not as extensive, particularly i terms of the use of inputs. This discrepancy means that the hypothesis could not be teste conclusively.

2. Small farms are defined as those up to two hectares Medium farms are between two an eight hectares in Kenya and between two and five hectares in Burkina Faso Farms of mor than eight hectares in Kenya and five hectares in Burkina Faso were considered large.

3. Only a small proportion of farmers became aware of extension messages through th radio. Virtually all who did considered the radio messages useful and suggested an increase i the length of the programs or a change to evening hours when they would not interfere wit farming activities.

4. Previous studies have used either an aggregate production function model or a producti vity decomposition model (Birkhaeuser, Evenson, and Feder 1991) To implement the latte an aggregate input index needs to be formulated, using cost shares as weights (Typically constant cost-share weights are derived from farm management studies) In contrast, an aggre gate production function specification does not require cost-share data and implicitly permi marginal products to differ from cost shares. This more flexible approach was used in th Kenya and Burkina Faso studies to estimate production coefficients using a simple Cob Douglas production function. More complex functional forms require more parameter est mates, and while they may enable more precise identification of parameters, they typical reduce the precision with which one can estimate the extension impact because of increase multicollinearity. National prices were used to aggregate the production of individual crop (Livestock production could not be included because of the lack of data)

5. We attempted to link the evidence on participation and adoption from the last section t the T&V impact on production, but the exercise was inconclusive for several reasons. First, th adoption of several practices was relatively incomplete in 1990. Second, the adoption of thes practices alone does not capture the full effects of extension on farm productivity. Good farm management includes proper timing and response to weather conditions and efficient alloca tion of household resources among activities. And there is considerable scope for managem

improvement after particular practices have been adopted. For instance, the practices themselves can be more effectively implemented after their initial adoption.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Bagchee, Aruna. 1994. *Agricultural Extension in Africa.* World Bank Discussion Paper 231. Washington, D. C.

Benor, Daniel, James Q. Harrison, and Michael Baxter. 1984. *Agricultural Extension: The Training and Visit System* Washington, D.C.: World Bank.

Bindlish, Vishva, and Robert E. Evenson. 1993 *Evaluation of the Performance of T&V Extension in Kenya* World Bank Technical Paper 208. Washington, D. C.

Bindlish, Vishva, Robert E. Evenson, and Mathurin Gbetibouo. 1993. *Evaluation of T&V-Based Extension in Burkina Faso.* World Bank Technical Paper 226. Washington, D.C.

Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder. 1991. "The Economic Impact of Agricultural Extension: A Review." *Economic Development and Cultural Change* 39(3):607–50

Cleaver, Kevin M 1993. *A Strategy to Develop Agriculture in Sub-Saharan Africa and a Focus for the World Bank* World Bank Technical Paper 203 Washington, D C.

Evenson, Robert E., and Yoav Kislev 1975 "Investment in Agricultural Research and Extension An International Survey." *Economic Development and Cultural Change* 23(3):507–21

Feder, Gershon, and Roger Slade 1986 "The Impact of Agricultural Extension: The Training and Visit System in India." *The World Bank Research Observer* 1(2) 139–61

Schorosch, Franz 1993. "The Agricultural Extension Initiative in Sub-Saharan Africa." World Bank, Africa Region, West Africa Department, Washington, D C Processed.

World Bank 1983 "Staff Appraisal Report. Kenya National Extension Project." Report 4426-KE Washington, D C Processed

Public and Private Agricultural Extension: Partners or Rivals?

Dina Umali-Deininger

This article examines the roles of the public and private sectors in agricultural extension. Extension services are classified according to their economic characteristics to identify areas where opportunities for private (for-profit and nonprofit) participation will arise. The author finds that commercialization of farm operations gives rise to demand for specialized client- and location-specific extension services that can be provided by private for-profit firms, although the main buyers will likely be market-oriented medium and large farmers. Because of market failures, some types of extension services will require public funding—although not necessarily public delivery. A critical government role in a pluralistic extension system would be to provide the appropriate regulatory framework to ensure fair competition and maintain quality standards.

Many countries recognize that ensuring an effective agricultural extension system is critical, especially in view of the major challenges facing the agricultural sector today. Rapidly growing populations have unleashed a spiraling demand for food, while the food-producing capacity in many nations is increasingly constrained both by diminishing opportunities to bring new land into production and by the declining productivity of over-cultivated areas caused by natural resource degradation (Crosson and Anderson 1992, Brown and Kane 1994, McCalla 1994). At the same time, the significant majority of the poor continue to depend on agriculture for most of their livelihood. Of the 720 million poor identified by the World Bank (1996a), 75 percent live in rural areas. Thus increasing farmers' incomes through improved productivity is an important element in agricultural development and poverty reduction strategies.

The adoption of improved technologies and sustainable farm management practices opens new opportunities to meet these challenges. The agricultural extension system is one of the primary vehicles for diffusing these technologies and therefore clearly has an important role to play in the development process.

The World Bank Research Observer, vol. 12, no. 2 (August 1997), pp. 203–24

The major dilemma for most governments, however, is who should sit in the driver's seat in the extension system—the public sector, the private for-profit sector, the private nonprofit sector, or some combination of the three.

The traditional view of the "public-good" character of agricultural extension services and the positive benefits they generate has led many governments to take exclusive responsibility for delivering extension services. A 113-country survey, conducted by the Food and Agriculture Organization in 1988–89, showed that national, state, or provincial governments conducted approximately 81 percent of the extension work in the surveyed countries (Swanson, Farmer and Bahal 1990). Nongovernmental organizations (NGOs) accounted for 7 percent; parastatals, 3 percent; universities, 2 percent; the private for-profit sector, 5 percent; and other providers, 2 percent.

Although selectivity bias and methodological problems have clouded the measurement of the benefits from agricultural extension, there is general agreement about its positive contribution to the adoption of new technologies and the increase in farm productivity and profits.[1] The debate centers more on the absolute magnitude of the returns. A review of forty-seven studies by Birkhaeuser, Evenson, and Feder (1991) reported "a significant and positive extension effect" in thirty-three cases. Eight studies that calculated net rates of return reported highly positive results in nearly all cases. The rates of return to extension varied across countries and commodities, ranging from 13 to 500 percent in Brazil, 75 to 90 percent in Paraguay, 100 to 110 percent in the United States and 14 to 15 percent in two states in India. Studies of extension returns in Asia, Africa, and Latin America showed returns of between 34 and 80 percent for nonstaple crops.

Three major developments have brought about a rethinking of the appropriate channel for delivering agricultural extension. First and most important, fiscal crises and economywide budget cutbacks, often associated with structural adjustment programs, have forced governments to make sharp reductions in the budgets of public extension programs. Financial sustainability and cost effectiveness have became the priority concerns.

Second, the poor performance of some public extension programs, as reflected by the slow adoption of extension messages, has spurred the search for alternative approaches to improve extension services. As Rivera (1991, p. [illegible]) notes, "Public sector extension [in the 1980s] was criticized for not doing enough, not doing it well, and for not being relevant." This failure was attributed to bureaucratic inefficiencies and the poor formulation and implementation of extension programs. As a result, many extension programs were inadequately funded and lacked a coherent link both with their farmer clientele and with the agricultural research sector. A recent review of thirty-one World Bank supported extension projects, of which 90 percent were modeled after the pub

lic training and visit extension system (T&V), found persistent problems arising from a failure to address the needs of particular farmers, inadequate human resource capacity, weak government commitment, and the likelihood that such programs will not be able to function without continued government support (table 1). These weaknesses in turn fueled the debate about the cost effectiveness of public extension systems. Although studies have shown the positive

Table 1. *Frequency of Problems Raised in Ex-Post Evaluation Reviews of Free-Standing Extension Projects*
(percent)

Description of problem	*Satisfactory projects n=23*	*Unsatisfactory projects n=8*	*Total projects n=31*
Funding			
Recurrent cost funding inhibiting field operations	87	100	28
Client orientation			
Inadequate research-extension linkage	74	100	26
Insufficient technology available	39	63	12
Entrenched "top-down" approach	48	75	17
Inadequate consideration for production risk and access to resources	39	50	13
Human resource capacity			
Limited practical knowledge of extension staff and technology available for dissemination	43	88	17
Low level of education limiting analytical and response capacity	22	38	8
Methodologies			
"Blue-print" approach (region, state, nation) ignored local characteristics and requirements	35	50	12
Inadequate attention given during preparation phase "Contact farmer" system was not effective, farmer group approach had better results	43	75	16
Government ownership			
Lack of government commitment	48	50	15
Weak monitoring or if developed not used	35	38	11
Sustainability of the system[a]			
Likely	33	n.a.	n.a.
Uncertain	58	60	n.a.
Unlikely	8	40	n.a.

Note: These ratings measure the extent to which a project met the relevant objectives in a cost-effective and administratively efficient manner.
a. Includes only twelve of the satisfactory and five of the unsatisfactory projects evaluated in 1989.
n.a. Not applicable.
Source: World Bank (1994a).

benefits from the T&V system (see Feder and Slade 1986; Bindlish and Evenson 1993; Bindlish, Evenson, and Gbetibouo 1993), it is by no means clear whether more cost-effective approaches and therefore higher rates of return could have been achieved through alternative nongovernment delivery systems. (The training and visit system is a purely public sector extension service delivery approach that rests on the use of "contact farmers" for the delivery of extension messages on a regular basis; see Benor and Baxter 1984.)

Third, agriculture's dependence on more specialized knowledge and technologies has changed the economic character of the services delivered by the extension system. The institutionalization of mechanisms that permit the seller to appropriate the returns from new inventions and new species of plants (for example, patents, copyrights, plant breeders' rights, and so on) has improved the private for-profit sector's incentives to provide extension services. The growing commercialization of agriculture and increased competition in domestic and international markets have further strengthened the economic incentives for farmers and other rural entrepreneurs to treat extension as another purchased input to agricultural production and marketing activities.

In the search for a new paradigm of the agricultural extension system, developing countries are wrestling with several questions: What are the appropriate roles for the public and private sectors? Can the private sector deliver services more efficiently? What are the welfare implications for small-scale farmers and the rural poor?

The Nature of Agricultural Information

Information transmitted to and from farmers through the agricultural extension system can be divided into two broad categories: pure information, and information that is embodied in new products or equipment (Ruttan 198[illegible]

Pure Agricultural Information

Pure information includes all types of self-standing advice on practices in four main areas:

- Cultural and production techniques, such as timing for planting and harvesting, use of inputs, animal husbandry and livestock health, crop protection, and farm-building design;
- Farm management, such as record-keeping, financial and organizational management, and legal issues;
- Marketing and processing information, such as prices, market options, storage procedures, packaging techniques, transport, and international standards for quality and purity; and

- Community development, such as the organization of farmers' associations.

Embodied Agricultural Information

Farmers can also obtain agricultural information indirectly, through technologies used in farm production, such as new agricultural equipment, chemicals, seeds, pharmaceuticals, and livestock breeds; technologies that facilitate farm management, such as telecommunications, laboratory equipment, computers, and software; and postharvest equipment, such as threshing, drying, milling, storage, and packaging technologies. Various combinations of these technologies have often been promoted as a package that includes credit and technical assistance. The most frequently cited examples are the Green Revolution technological packages of high-yielding crop varieties (rice or wheat), irrigation, credit, fertilizers, pesticides, and extension.

The potential providers of agricultural extension services fall into three major groups: the public sector, the private nonprofit sector, and the private for-profit sector. The distinction is important because of the different range of services each, and particularly the private for-profit sector, has the incentive to deliver. The public sector is represented by ministries or departments of agriculture at the federal, state, provincial, and municipal levels and by international and regional organizations, such as the Consultative Group on International Agricultural Research and the South East Asian Research Center for Agriculture. The private for-profit sector in the extension system comprises all economic agents whose objective is to generate profits directly or indirectly for their owners, members, or shareholders. This sector covers commercial enterprises operated by a farmer or group of farmers, including cooperatives; commercial production and marketing firms, such as input manufacturers and distributors, agromarketing and processing firms; commodity boards; and private consulting and media companies (publishing and telecommunications firms). The private nonprofit sector differs from the for-profit sector in one important respect: rather than distributing the residual earnings (if any) to individuals who exercise control, it must reinvest profits to finance future activities. In this category are nongovernmental organizations (NGOs), universities, foundations, professional and trade associations, and other noncommercial groups.

Economic Incentives for Delivering Extension Services

Extension information could be classified according to its economic character, based on whether it is closer to being a *public good* or a *private good*, using the

economic principles of rivalry and excludability. Rivalry (or subtractability) applies when one person's use or consumption of a good or service reduces the supply available to others. The purchase of an improved small-scale hand tractor reduces the total supply of equipment available to others. Excludability applies when only those who have paid for the product or service benefit from it (Feldman 1980, Kessides 1993). The farmer who purchased the hand tractor has the sole proprietorial right to use the equipment. Any good both rival and excludable is a private good; those that are neither rival nor excludable are public goods. Private firms are unwilling to supply goods and services with public-good characteristics because restricting the benefits derived from those goods only to those who paid for them is usually impossible. A farmer will not be willing to pay for information on soil conservation techniques that is also reported by radio, because other farmers can freely tune in and obtain the same information.

Purely public and purely private goods occupy opposite ends of the economic spectrum. In between the two extremes are *toll goods* and *common-pool goods* (figure 1). Toll goods are excludable, but not rival; for example, the supply of information provided by a private extension consultant exclusively to a group of farmers is not reduced by the addition of another member to the group. Toll goods cannot be accessed by people who do not pay for the service although a decision of a member farmer to pass on information that the group had paid for would not reduce the group's consumption. The ability to exclude those who have not paid for the service provides the incentive for the private sector to supply such goods. Public regulation, however, will be necessary to establish property rights, conditions of competition, and pricing and quality standards for the toll services.

Common-pool goods are those that are rival but not excludable; in other words other people cannot be stopped from using them. For example, the purchase of high-yielding self-pollinated seeds such as rice and wheat reduces the supply of such seeds, but their ease of replicability makes exclusion difficult and costly in the long run. Farmers do not buy rice and wheat seeds every season, because they can set aside part of their harvest for planting the next crop.

Some types of information embodied in technologies produce *externalities*. These occur when an individual, in using (or producing) a product or service for which payment is received (or made), coincidentally benefits other people from whom payment cannot be exacted or adversely affects others whose claims for compensation cannot be enforced (Pigou 1932). Typically, the individual responsible for the externality will not consider the positive or negative effects when deciding how much to produce or consume. As a result, either too little (for positive externalities) or too much (for negative externalities) is produced or consumed. Markets may develop their own mechanisms to account for these

Figure 1. *Economic Classification of Agricultural Information and Technologies Delivered by the Agricultural Extension System*

Rivalry	Excludability: Low	Excludability: High
Low	**Public goods** • Nonexcludable agricultural information (LT) • Mass communication of agricultural information	**Toll goods** • Nonexcludable agricultural information (ST) • Excludable agricultural information (Cultural and production practices, farm management, marketing, processing)
High	**Common-pool goods** • Modern technologies (Self-pollinated seeds [LT])	**Private goods** • Modern technologies (Machinery, chemicals,* hybrid seeds, self-pollinated seeds (ST), biotechnology products,* veterinary supplies and pharmaceuticals)

*Items may involve externalities. ST = short term; LT = long term.
Source: Umali and Schwartz (1994).

externalities. If these mechanisms do not develop naturally, government intervention of some kind is justified, such as subsidizing the desirable activity to increase consumption, or regulating or taxing the use of those activities that result in a negative externality to reduce private consumption to socially desirable levels.

Problems of asymmetric information also arise with the use of some types of agricultural information or technologies. Consumers may be unable to assess the quality of the product they are purchasing—for example, the difference between good or bad seed, an adulterated or pure livestock feed, or good or bad advice. In some cases, suppliers may enforce quality control to maintain brand loyalty (Blankart 1987), or trade associations may enforce industrywide quality standards. Even if these market-based measures exist, however, the public sector should be responsible for maintaining and enforcing quality standards.

These concepts provide the basis for classifying the type of information delivered by agricultural extension providers and thus the appropriate roles for the public and the private sectors.

Agricultural Technologies

For most modern technologies, such as agricultural machinery, agricultural chemicals, hybrid seeds and livestock, and veterinary supplies and pharmaceu-

ticals, agricultural information is embodied in the invention. These technologies are classified as private goods because of their high subtractability and excludability. The use of legal mechanisms (such as patents, copyrights, and plant variety protection) provides a high degree of excludability. These technologies are thus especially attractive to producers and distributors, which have the incentive to supply these technologies and the supporting technical information to all farmers at socially optimal levels. Where these products or services are too costly for individual farmers, channeling them through client organizations (farmer associations or commodity foundations) enables poor farmers to take advantage of new opportunities. These institutions, in consultation with their members, determine the level of services needed and charge farmer members according to the services provided.

The use of some technologies involves externalities or spillover effects. A farmer's use of a veterinary drug or vaccine results in positive externalities because it reduces the risk that livestock disease will be transmitted to neighboring animals. Inappropriate use of a pesticide can result in negative externalities polluting the food supply, causing pest-resistance, and destroying natural enemies. *Different mechanisms have evolved in the market to deal with these* externalities. Increasing concern about pesticide use has spurred the search for new technologies to detect pesticide residues in food and the effects of exposure to toxic materials and has changed consumer preferences, as reflected by the price premiums consumers are willing to pay for organically grown, pesticide free produce (Phipps 1989).

Some types of agricultural technologies, such as self-pollinated seeds, are private goods only in the short term, that is, in the introductory phase. In the medium- to long-term, they become common-pool goods. Private firms supplying these seeds, therefore, have to compete with farmers who use their own seeds, resulting in very low or negligible profit margins. Consequently only relatively small local private firms with low overhead costs that produce consistently high-quality seeds may be expected to earn adequate profits (Jaffee and Srivastava 1992). The size of the market and the efficiency of the operation determine whether the private sector will participate in the production and distribution of self-pollinated seeds. If these conditions are not met, state and local governments and private nonprofit agencies will have to distribute the seeds

General or Nonexcludable Information

Information designed to improve existing cultural and production practices farm management, or marketing and processing techniques and provided by traditional agricultural extension approaches is a toll good in the short term Although transmitting such information does not reduce its availability to oth

ers, attendance at extension meetings is constrained by the facilities and infrastructure and by the time it takes for word of mouth to reach a critical mass of interested farmers. But the diffusive nature of nonexcludable information transforms it into a public good quickly. Similarly, nonexcludable information that is transmitted through public broadcasts and public distribution is a public good. Thus, how quickly information is diffused determines whether the private sector has an incentive to provide it. If the information diffuses easily, the possibilities of charging for it are limited, and private firms will have little or no incentive to provide such services; thus delivery of nonexcludable information will remain the responsibility of the public sector or of private nonprofit agencies.

In special situations, the private for-profit sector will provide nonexcludable information. For example, agroprocessing and marketing firms will undertake extension activities when the revenues they realize from a more assured supply, improved timing, and higher quality are greater than their costs of providing the extension information. Private firms may also provide information to their customers about the appropriateness or the range of uses of their products as part of their advertising campaigns to expand or protect their market shares.

Excludable or Specialized Information

As farm operations become more commercialized and agricultural technology more specialized, the corresponding extension services needed to support these activities also become highly specialized. Such specialization lends exclusivity to the information and, therefore, the extension activity. For example, the results of a soil analysis or the development of computer programs to facilitate farm operations are location- and client-specific. Such information may not be useful to other farmers, and even if it is, the farmers receiving the information may reserve it for their own exclusive use and not transmit it freely to others, thus slowing or blocking the diffusion process. These characteristics qualify specialized information as toll goods and provide adequate economic incentives for the private for-profit sector to supply such information at optimal levels. Asymmetric information problems, however, increase the difficulty of assuring quality. Unless the private fee-for-service extension industry can effectively police itself to ensure the quality of the information communicated, public intervention will be necessary to enforce quality standards and legal contracts.

In some cases, significant externalities may also be associated with specialized types of information. The prime example is natural resource management techniques, which directly benefit the farmer-cultivator by promoting sustainable agricultural production and at the same time benefit society—present and fu-

ture. In these instances, public-sector interventions, such as subsidies to promote beneficial techniques that could raise usage to socially optimal levels, are needed. The private nonprofit sector can also play an important role in disseminating this information.

Private Agricultural Extension Services

Experience in several countries attests to the broad scope of agricultural extension services that the private sector could deliver efficiently and profitably. These are described below.

Agribusiness Enterprises

Agroprocessing and marketing firms promote a wide range of commodities through the use of in-house agricultural extension services for farmers who grow for them under contract arrangements. These private firms typically focus on the type and levels of use of inputs, disease prevention or control, and harvest and postharvest techniques. They are designed to increase farmer output, reduce postharvest losses, and improve the quality, consistency, and timeliness of the crop. Examples from various developing countries include broilers, fruit and vegetables, cotton, and tobacco (Zijp 1991, Rama 1985; de Janvry, Runsten and Sadoulet 1987; Rogers 1987; Glover 1990; Schwartz 1992; Okidegbe 1990; Venkatesan 1995; and Agribusiness Worldwide 1982). In Kenya, the large local fresh produce exporters, such as Indufarm, Sunripe, MHI, and Homegrown, currently provide inputs and extension services to farmers producing for them on contract.

During the 1970s dairy farmers in Argentina faced serious obstacles. Livestock there was unproductive; the milk supply was unstable and often of poor quality. These problems were mainly the result of poor animal nutrition and inadequate farm hygiene. The two largest dairy processors, Santa Fe–Cordoba United Cooperatives (SANCOR) and La Serenisima, whose own growth was jeopardized by the plight of the dairy farmers, launched extension programs to overcome these constraints. SANCOR created an extension department with eight regional offices, each managed by an agronomist assisted by middle-level technicians. Each office provided extension services to almost forty cooperatives and assisted small groups of farmers (usually six to fifteen) who met monthly to discuss a visited farm's progress and problems. SANCOR initially financed technical assistance for these small groups, but after thirty months, each group took on the cost of the professional agronomist. By 1990 SANCOR had 120 farmer groups participating in the program. Artificial insemination services as well as

accelerated heifer-rearing programs were also organized. In addition, SANCOR supplied and financed a varied list of farm inputs. As a result of the program, milk production increased by 15 percent between 1976 and 1985 despite a 24 percent decrease in the number of dairy farms participating in the cooperative.

La Serenisima also formed a strong technical assistance department, but it deliberately targeted medium- to large-scale farmers. It established five regional offices, each with five branch offices, which worked with groups of up to twenty-five farmers. Magazines, bulletins, radio broadcasts, and television programs were all part of the development effort. During 1978–85, although the dairy farm areas feeding La Serenisima shrank by more than 6 percent, production increased by almost 50 percent (World Bank 1989).

Private Consulting Firms

Private consulting firms provide agricultural extension services in many developing countries (Pray and Echeverría 1990; World Bank 1987, 1990a, 1990b). In 1988, for example, more than 2,000 private consulting firms were working in Brazil, largely catering to the highly specialized extension needs of the commercial livestock sector. Firms conducting extension services in Asia tend to concentrate on plantation crops, and a growing number are staffed by former plantation managers and technicians.

Farmer Associations

A large number of farmer associations also provide agricultural extension services to their members. These associations cover a wide range of commodity-specific topics, including new technologies, production techniques, farm management, disease prevention and control, and marketing and processing procedures.[1]

In France large cooperatives hire technical advisers directly and cover the cost by charging members a fee. Four major French farm organizations also manage agricultural development services at the local, regional, and national level including seventeen applied research institutes. The National Fund for Agricultural Development is a joint venture under the control of the *Association Nationale pour le Developpement Agricole* (National Association for Agricultural Development), an association subject to government control and jointly managed by representatives of the government and farm organizations. This fund accounts for approximately one-fourth of the total resources and is also used to ensure that each area receives its fair share of funding (LeGouis 1991).

A similar association (AACREA) in Argentina, based on the French model, consists of 15 regional groups and 176 local groups with more than 2,000 mem-

bers. Each local group of eight to twelve farmers is led by a professional agronomist who visits each of the farmers in the group for one day a month. The farmers also visit a member's farm each month to discuss specific operations and problems. The cost to each farmer is about $60 a month (equivalent to one farm laborer's monthly wage), of which 80 percent covers the professional's fee and the rest goes to the organization.

In the Central African Republic, the National Federation of Central African Livestock Farmers (FNEC) was organized to facilitate the distribution of veterinary drugs after the government livestock service nearly collapsed in 1973. In 1989 FNEC also began providing extension services and education programs to its members, who include more than 60 percent of all herders in the country (Umali and Schwartz 1994).

In Zimbabwe 9,000 small-scale commercial farmers and 65,000 smallholders belong to the Zimbabwe Farmers Union. The Commercial Farmers Union (CFU) has a membership of 4,450 large-scale farmers and serves as the umbrella organization for several commodity-specific associations, such as those for flower and tobacco farmers, coffee growers, and sheep and cattle farmers. The CFU is the organizing force behind a range of research and extension activities, such as the Agricultural Research Trust, a weekly magazine called *The Farmer*, and bimonthly reports. In addition, the CFU provides leadership for some commodity specific activities, such as the Cotton Training Center at Kadoma (Schwartz 1992).

Several factors stimulated the farmers' associations in Zimbabwe to get involved in research and extension. First, the emphasis of public extension shifted from a plantation to a smallholder clientele. Second, the shift in emphasis left commercial farmers anxious to ensure their access to quality services. Third, a number of European extension staff who left public service became available for the associations to hire. The associations employ extension specialists (referred to as district councillors) who work directly with the farmers in an area, usually on specific commodities. Most farmers are members of multiple associations and have a variety of information sources (including private consultants). The CFU and the commodity-specific organizations maintain links with and provide some services to public-sector research and extension. The CFU extension staff provides training at the Cotton Training Center and engages in collaborative efforts in the field such as the Agricultural Research Trust farm trials in communal areas and pesticide demonstrations at farm shows (Schwartz 1992).

Nongovernmental Organizations

NGOs have also began to assume a greater role in agricultural extension, frequently focusing on areas that the government has neglected. One reason for

their success has been their community-based focus. In West Africa, for example, the *Se Servir de la Saison Sèche en Savanne et en Sahel* (the 6-S Program for the Savannah and the Sahel) promotes village organizations, helps groups establish community development programs, and provides funding and technical assistance for projects including village crafts, cereal banks, market gardening, soil conservation, and reforestation. With an annual budget of $1.25 million, 6-S is now operating in Burkina Faso, Mali, and Senegal. Since its founding in 1976, it has established 2,000 farmer organizations (averaging eighty members per group) in about 1,000 villages (Amanor and Farrington 1991). In Northern Ghana the Agricultural Information Service, funded by the Presbyterian Agricultural Station at Langbensi, works with more than twenty church-based agricultural stations and coordinates with the government research station at Nyankpala (Amanor and Farrington 1991). In Pakistan the Aga Khan Foundation is involved in rural development projects, dealing with technology transfer and training of farmer representatives to become specialists in livestock, plant protection, marketing, and forests (Khan 1992).

Some NGOs have organized regional networks. In Latin America, eight NGOs representing seven countries have formed the *Consorcio Latinamericano Agroecologa y Desarollo* (CLADES — Latin American Consortium on Agroecology and Development) to promote, develop, and diffuse agricultural information to small-scale farmers, in association with institutions committed to sustainable agriculture (Altieri and Yurjevic 1989).

Many NGOs perform both research and extension activities. In India the Bharatiya Agro-Industries Foundation (BAIF) serves 8,000 villages in six states with a program to improve dairy cattle. It operates 450 artificial insemination centers and provides veterinary and extension services (Amanor and Farrington 1991). In fact, some state governments have subcontracted with BAIF to provide artificial insemination services (World Bank 1996b).

Fee-for-Service Extension

The demand for agricultural extension services, and therefore the willingness to pay for them, depends on the expected benefits from the new information. The type and level of demand will be determined by the magnitude and value of the farm household's marketable surplus, the cost of the extension service, and the additional income generated as a result of adopting the extension information. Given that a farmer cannot buy only part of the service and that a fixed or negotiated fee is usually paid, medium- and large-scale producers can spread the cost, resulting in lower per-unit costs and higher rates of return. Conse-

quently, the larger the farm operations, the greater the potential demand for "fee-for-service" extension.

Strong market competition associated with high-value tradable commodities could also enhance effective demand for new information, because the new information and technologies may provide the competitive edge. In contrast, because the value of their marketable output is low, resulting in higher per-unit costs, small-scale farmers typically find it less attractive or profitable to "purchase" the extension service. Subsistence farmers have limited, if any, incentive to pay for extension services.

Government policies can greatly affect the demand for extension services, through their (direct and indirect) influence on commodity prices and aggregate demand. High (direct and indirect) taxes on agriculture reduce farmers incentives to adopt improved technologies. Similarly, high inflation and macroeconomic instability limit access to the capital necessary to finance such new technologies. The allocation and level of public expenditure on rural roads, markets, and irrigation infrastructure, for example, influence the development potential of particular localities and thus the return on investments in technologies that enhance productivity. Public expenditures on education, especially in rural areas, have a strong influence on the capacity of farmers and consumers to absorb new information.

A major implication of the shift in the classification of information from a "free good" to a "purchased good" is that the demand for paid agricultural extension services will originate almost exclusively from market-oriented farming operations and particularly from medium- and large-scale farmers. It also follows that such farmers will be more capable of sustaining a "fee-for-service" agricultural extension business. Conversely, private for-profit firms will tend to neglect areas composed of more marginal farmers.

Private Supply

The costs of and returns to the services a firm provides will determine the level at which private extension services will be supplied. Costs are influenced by the supply of and demand for qualified extension personnel, economies of scale in delivery operations, and government policies; revenues reflect the factors that influence farmer demand. The use of and returns to extension subsequently affect the nature and level of supply of domestic information (pure or embodied in technologies). Government restrictions on imports of technology (such as improved seeds, livestock breeds, and agricultural chemicals) reduce the menu of products available for extension providers to disseminate.

Because profitability is the main criterion for private extension providers, their tendency is to cultivate a clientele of commercial farmers. The issue of

"cherry-picking" thus becomes a major concern. The Chilean experience illustrates this problem. When Chile privatized its extension system in the 1970s, commercial farmers were not seriously affected, but small-scale and subsistence farmers were priced out of the extension market. To remedy this inequity, the government had to take active measures to target extension services to these farmers.

Provided farmers can overcome the difficulties of organizing into a group, farmers' associations can allow small farmers to pool their resources to purchase extension information that individual farmers may not be able to afford on their own. The association may employ part-time or full-time staff or contract out its extension requirements to consulting firms.

When should extension be funded by the public sector? Where extension delivers public goods and information with high externalities, such as environmental or conservation-related information, complete privatization is neither desirable nor feasible. Two other arguments could justify public subsidization of extension to small farmers: first, when small farmers may be unaware of the benefits of improved technologies and unable to afford them, and second, when small subsistence farmers may derive considerable nonmonetary benefits (including better nutrition and health) from adopting new technology.

Public-Private Partnerships

Several Latin American countries are attempting to address problems with public extension systems by integrating the private sector into public extension activities. These new approaches include subcontracting to the private sector and an extension voucher system, both of which have partial cost-recovery components.

Private Sector Subcontracting

Subcontracting extension delivery to the private sector (profit and nonprofit) could provide a mechanism for getting around the institutional inefficiencies associated with public delivery (table 2). In 1992 Chile launched an extension project that included subcontracting extension services to private consulting firms (Wilson 1991). To qualify as subcontractors under the plan, private firms must meet technical and professional staffing criteria, bid for contracts, and agree to have their activities monitored by a designated public agency. Similar programs have also been launched in Mexico and Venezuela (World Bank 1994b, 1995b).

Table 2. *New Approaches to Agricultural Extension Delivery*

	Funding		Delivery	
Country	*Public*	*Farmer cost sharing*	*Public*	*Private*
Chile				
Large farmers	n.a.	Full costs	Commercial	Commercial
Small/subsistence farmers	Majority	Increasing share	Subcontracted	Subcontracted
Colombia	National trust fund + municipality	Free	Municipality	
Mexico	National trust fund	Increasing share (FG)	n.a.	Subcontracted
Venezuela	National + state + municipality	Increasing share (FG)	n.a.	Subcontracted
Nicaragua	National	Increasing share (FG)	(Vouchers)	(Vouchers)

n.a. Not applicable.
FG Farmer group formation
Source: Wilson (1991), World Bank (1993, 1994b, 1995a, 1995b)

Extension Vouchers

In an on-going pilot program in Nicaragua, the government issues extension vouchers to farmers allowing them to choose their extension supplier, either public or private. Suppliers include established input providers, nongovernmental organizations, and former personnel of the *Instituto Nicaragense de Reforma Agraria* (Nicaraguan Institute for Agrarian Reform) and the Ministry of Agriculture who have established private firms. Under the program, farmers form small groups of at least ten members who submit a signed master service contract to the Ministry of Agriculture for review. A three-member panel, consisting of a representative from the agriculture ministry, the Institute of Agricultural Technology, and the National Development Bank (BANADES) award the contracts according to established criteria. Each farmer group receives fifteen vouchers a year, 40 percent of which can be applied to group visits and the remaining 60 percent to individual visits. Each voucher was valued at $15 and can be redeemed only by qualified providers at the regional branches of BANADES (World Bank 1993; Keynan, Olin, and Dinar 1997).

The Nicaraguan government has implemented several measures to ensure the effectiveness and quality of the services provided, including the accreditation of technical assistance providers by the Ministry of Agriculture, the establishment of selection criteria for these providers, and the development of standards for monitoring and evaluating the program. Initially, a promotional campaign was launched to announce and describe the program to potential

suppliers and clients. Technical assistance providers can be either individuals or legally established organizations, such as NGOs, service cooperatives, and private firms. Some general criteria include at least three years experience at the professional level (university graduate) in technology generation and transfer of technology to small- and medium-scale farmers; at least five years experience at the technical level in these areas; satisfactory completion of the basic training course offered by NIAT; and formal enrollment with the Ministry of Agriculture. To ensure the integrity of the suppliers, independent auditors review the performance of the technical assistance providers.

Strengthening Client Orientation

Three approaches introduced in Latin America are designed to foster increased client orientation: farmer cofinancing of extension services, decentralization of financing and planning to municipalities, and use of farmer groups. Under the first approach, a user fee is charged to cover part of the cost of the service, to foster a more demand-driven system, and to serve as an important transitional phase in developing a market for fee-for-service extension. The minimal charge is designed to increase farmers' commitment to extension advice and to empower them as fee-paying customers. Cost-recovery components are incorporated into programs in Chile, Mexico, Nicaragua, and Venezuela. (For a description of the program in Nicaragua, see the article by Keynan, Olin, and Dinar in this volume.)

Under the Integrated Technology Transfer Program in Chile, for example, farmers are required to finance 15 percent of the extension cost (Wilson 1991). In Mexico the government plans to raise the level of cost-sharing in some of the more commercial areas to 100 percent over time (World Bank 1994b). The programs in Colombia and Venezuela feature cost-sharing by municipalities as part of the government's fiscal decentralization program (World Bank 1995a, 1995b). Decentralization of control to the municipalities is also aimed at improving client orientation and ownership. Increased reliance is being placed on farmer groups as a primary point of contact between farmers and extension providers. This strategy is intended not only as a way to recover some of the costs of extension, but also to capitalize on the economies of scale associated with extension delivery.

The extent to which the participants in a program can be induced to adopt and consistently maintain sound new technologies is a vital consideration. A review of Chilean experience revealed that after three years, both farmers' and private consultants' interest in the program declined. Several factors seem to have been involved. First, demand from farmers declined as the effective cost of the extension services increased—the farmers' share of the cost increased while

the package of services provided was reduced. Second, the quality of the exte sion services deteriorated. Third, a seemingly "top-down" approach reduc farmers' sense of ownership. This was due, in part, to poor feedback mech nisms in the system; because new demands were not transmitted back to t program planners and the advice was not tailored to local needs (McMah 1994).

Social and economic factors strongly influence the change in demand th occurs in response to a change in the price of extension services. Considerati of these factors is essential in predicting and "managing" farmers' response changes in the fee structure, and extension staff need to be trained to be mc responsive and flexible to meet the changing extension demands of farmers.

Conclusion

In view of the challenges facing farmers today, the structural transformation agricultural extension systems around the world is unavoidable. This change being hastened by the significant fiscal constraints faced by many countri Several countries are actively pursuing greater involvement by the private no profit and for-profit sectors as a means of improving the effectiveness ar sustainability of agricultural extension systems. Capitalizing on the compar tive advantage of each of the different sectors helps to ensure the success of th endeavor.

For countries that are at the crossroads, fostering a smooth transition fro the traditional purely public agricultural extension system to a more pluralist one will require close attention to four key areas: source of funding, client ta geting, cost recovery, and delivery channels. The first major issue is who w pay for the extension service—the government, farmers, or private nonpro organizations? The responsibility for funding different types of extension prod ucts will be influenced primarily by the economic characteristics of the exte sion information to be disseminated and the structure of the local farm secto Farmers will pay willingly only for private goods or information that is chara terized as a toll good.

The second and third major concerns—client targeting and cost recovery are closely associated with public sector financing. Budgetary constraints hav brought to the forefront the issue of targeted as opposed to blanket provision extension services. Given the government's limited resources, should the publ sector focus its efforts exclusively on providing an "extension safety net" fo small and subsistence farmers and let medium- and larger-scale farmers rely o private providers? And should the public sector institute cost recovery? Charg ing farmers a nominal sum for services can encourage them to exercise the

rights as information consumers, thereby ensuring program effectiveness. The business relationship introduced with cost recovery also imposes new demands on extension providers. In particular, it implies that farmers can choose who will provide the service and that the extension providers must respond quickly to meet the farmers' information demands. Consequently, marketing and sales skills have to be included in the basket of technical skills that extension agents offer.

Whether delivery of extension services should remain a public monopoly or be fully or partially subcontracted to private organizations, either for-profit or nonprofit, will also need to be resolved. Two determinants of the optimal choice of institutional delivery channel will be the capacity of the contractor to minimize production costs and the capacity of the government to minimize the transaction costs associated with monitoring and regulating services. Another critical factor is the capacity of the domestic private sector. In some countries, the development of private extension capacity has been suppressed, and it is absent or very weak. Special programs that encourage public extension agents to leave public service and set up private operations have been used in Latin America to address this problem. Like an infant industry, the development of for-profit and nonprofit sectors may have to be promoted in the initial stages. The public sector should develop an appropriate regulatory framework with explicit standards and monitoring operations to ensure that quality standards are met.

Promoting a more pluralistic agricultural extension system requires important ideological adjustments on the part of governments and international donors. This new perspective is crucial for agricultural and economic progress.

Notes

Dina Umali-Deininger is an economist with the Rural Development Management Unit, South Asia Region, of the World Bank. Previous drafts of this paper greatly benefited from discussions with and comments by Klaus Deininger, Gershon Feder, Matthew McMahon, Willem Zijp, participants at seminars at the World Bank, and anonymous reviewers of this journal.

1. Selectivity bias occurs because measurements of the economic rates of return of extension are based on country- and commodity-specific studies. Thus, these measurements may be subject to a systematic upward bias relative to the "global average" because they do not account for less successful extension activities in other commodities and countries. Methodological problems include an upward bias that occurs because more innovative farmers tend to adopt extension advice first; a downward bias that results when farmers share information among themselves rather than receiving it directly from the extension service; the influence of location-specific variables, the appropriate designation of the extension variable; and times of measurement (early or late stages). Exogenous factors, such as price, trade, education, technology, and monetary and fiscal policies, could also influence the results.

2. Some successful associations include *Asociacion Argentina de Consorcios Regionales de Experimentacion* (AACREA—Argentine Association of Regional Agricultural Experimentation Consortia) in Argentina (Tobar 1996), the Tobacco Producers', Commercial Cotton Growers', Commercial Grain Producers', Ostrich Producers, and Crocodile Producers' associations in Zimbabwe (Schwartz 1992); the *Cooperativa Integral Campesina* (COINCA—Integrated Farmers Cooperative) for grapes in Bolivia (Tendler 1983); the Anand Milk Union Ltd. and other dairy marketing cooperatives under Operation Flood in India (Umali and others 1994); the Vegetable Cooperative in Uganda (Narayanan 1991); and the Poultry, Flower, and Vegetable associations in Turkey (Zijp 1991). COINCA was instrumental in teaching farmers techniques for improving the quality of their grape output, thus enabling them to take advantage of the quality premiums offered by the distilleries.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

The Agribusiness Worldwide. 1982. November/December:19–26.

Altieri, Miguel A., and David Yurjevic. 1989. "The Latin American Consortium on Agroecology and Development: A New Institutional Arrangement to Foster Sustainable Agriculture Among Resource-Poor Farmers." *Development Anthropology Network* 7(1):1[illegible]–19.

Amanor, Kojo, and John Farrington. 1991. "NGOs and Agricultural Technology Development." In Rivera and Gustafson.

Benor, Daniel, and Michael Baxter. 1984. *Training and Visit System.* Washington, D.C.: World Bank.

Bindlish, Vishva, and Robert E. Evenson. 1993. *Evaluation of the Performance of [illegible] Extension in Kenya.* World Bank Technical Paper 208. Washington, D.C.

Bindlish, Vishva, Robert E. Evenson, and Mathurin Gbetibouo. 1993. *Evaluation of [illegible] Based Extension in Burkina Faso.* World Bank Technical Paper 226. Washington, D.C.

Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder. 1991. "The Economic Impact of Agricultural Extension: A Review." *Economic Development and Cultural [illegible]* 39(April):607–50.

Blankart, Charles B. 1987. "Limits to Privatization." *European Economic Review* 31:3[illegible].

Brown, Lester R., and Hal Kane. 1994. *Full House: Reassessing the Earth's Population Carrying Capacity.* Worldwatch Environmental Alert Series. New York: W. W. Norton and Company.

Crosson, Pierre, and Jock R. Anderson. 1992. *Resources and Global Food Prospects: Supply and Demand for Cereals to 2030.* World Bank Technical Paper 184. Washington, D.C.

de Janvry, Alain, David Runsten, and Elisabeth Sadoulet. 1987. "Technological Innovation in Latin American Agriculture." Program Papers Series 4. Inter-American Institute for Cooperation on Agriculture. San Jose, Costa Rica. Processed.

Feder, Gershon, and Roger Slade. 1986. "The Impact of Agricultural Extension: The Training and Visit System in India." *The World Bank Research Observer* 1(2):139–61.

Feldman, Allan M. 1980. *Welfare Economics and Social Choice Theory.* Boston: Kluwer Nijhoff Publishing.

Glover, David. 1990. "Contract Farming and Outgrower Schemes in East and Southern Africa." *Journal of Agricultural Economics* 41:303–15.

Jaffee, Steven, and Jitendra Srivastava. 1992. *Seed Systems Development: The Appropriate Roles of the Private and Public Sectors.* World Bank Discussion Paper 167. Washington, D.C.

Kessides, Christine. 1993. *Institutional Options of the Provision of Infrastructure.* World Bank Discussion Paper 212. Washington, D.C.

Keynan, Gabriel, Manuel Olin, and Ariel Dinar. 1997. "Cofinanced Public Extension in Nicaragua." *The World Bank Research Observer* 12(2): 225–47.

Khan, Shoaib S. 1992. "Nongovernmental Organization Alternatives and Fresh Initiatives in Extension: The Aga Khan Rural Support Programme Experience." In Jock R. Anderson and Cornelis de Haan, eds., *Public and Private Roles in Agricultural Development: Proceedings of the Twelfth Agricultural Sector Symposium.* Washington, D.C.: The World Bank.

LeGouis, Michel. 1991. "Alternative Financing of Agricultural Extension: Recent Trends and Implications for the Future." In Rivera and Gustafson.

McMahon, Matthew. 1994. Personal communication. World Bank, Country Department III, Latin America and the Caribbean Region, Washington, D.C.

McCalla, Alexander F. 1994. "Agriculture and Food Needs to 2025: Why We Should Be Concerned." Sir John Crawford Memorial Lecture, Consultative Group on International Agricultural Research. International Centers Week, October 27, 1994. CGIAR, Washington, D.C. Processed.

Narayanan, Anandan. 1991. "Enhancing Farmers' Income through Extension for Agricultural Marketing." In Rivera and Gustafson.

Okidegbe, Nwanze. 1996. Personal communication. World Bank, Agriculture and Natural Resources Department, Washington, D.C.

Phipps, Tim T. 1989. "Externalities and the Returns to Agricultural Research: Discussion." *American Journal of Agricultural Economics* 71(2): 466–67.

Pigou, Arthur C. 1932. *The Economics of Welfare.* 4th ed. London: Macmillan.

Pray, Carl, and Ruben Echeverria. 1990. "Private Sector Agricultural Research and Technology Transfer Links in Developing Countries." In David Kaimowitz, ed., *Making the Link: Agricultural Research and Technology Transfer in Developing Countries.* San Francisco: Westview Press.

Rama, Ruth. 1985. "Do Transnational Agribusiness Firms Encourage the Agriculture of Developing Countries? The Mexican Experience." *Food Systems* 37(3): 331–43.

Rivera, William M. 1991. "Agricultural Extension Worldwide: A Critical Turning Point." In Rivera and Gustafson.

Rivera, William M., and Daniel J. Gustafson, eds. 1991. *Agricultural Extension: Worldwide Institutional Evolution and Forces for Change.* London: Elsevier.

Rogers, William L. 1987. "The Private Sector: Its Extension Systems and Public/Private Coordination." In William M. Rivera and Susan G. Schram, eds., *Agricultural Extension Worldwide: Issues, Practices, and Emerging Priorities.* New York: Croom Helm.

Ruttan, Vernon. 1987. "Some Concerns about Agricultural Research Policy in the United States." *Quarterly Journal of International Agriculture* 26(4): 145–56.

Schwartz, Lisa. 1992. "Private Technology Transfer in Sub-Saharan Africa." Tropical Research and Development, Inc. Washington, D.C. Processed.

Swanson, Burt, Bertram J. Farmer, and Ram Bahal. 1990. "The Current Status of Agricultural Extension Worldwide." In Food and Agriculture Organization, *Report of the Global Consultation on Agricultural Extension.* Rome.

Tendler, Judith. 1983. "What to Think About Cooperatives: A Guide From Bolivia." The Inter-American Foundation, Rosslyn, Virginia. Processed.

Tobar, Julio G. 1996. "CREA-AACREA, Associative Private Agricultural Extension in Argentina." Case study prepared for the Workshop on Alternative Mechanisms for Funding and Delivering Extension, held on June 18–19, 1996. World Bank, Agriculture and Natural Resources Department, Washington, D.C. Processed.

Umali, Dina L., Gershon Feder, and Cornelis de Haan. 1994. "Animal Health Services: Finding the Balance between Public and Private Delivery." *The World Bank Research Observer* 9(1):71–96.

Umali, Dina L., and Lisa Schwartz. 1994. *Public and Private Agricultural Extension: Beyond Traditional Frontiers.* World Bank Discussion Paper 236. Washington, D.C.

Venkatesan, Venkatachalam. 1995. "The World Bank's Experience with T&V in Africa: Issues for the Future." In *Agricultural Extension in Africa: Proceedings of an International Workshop.* Wageningen, Netherlands: Technical Centre for Agricultural and Rural Cooperation.

Wilson, Mark. 1991. "Reducing the Costs of Public Extension Services: Initiatives in Latin America." In Rivera and Gustafson.

World Bank. 1987. "Mexico Agricultural Extension Project (PROCATI)" Report 6620-ME. Country Department II, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1989. "Argentina Agricultural Sector Review, Volume II: Technical Annex." Report 7733-AR. Country Department I, Latin America and the Caribbean Region, Washington D.C. Processed.

———. 1990a. "Agricultural Extension: The Next Step." Policy and Research Series 13. Agriculture and Natural Resources Department, Washington D.C. Processed

———. 1990b. "Brazil, Key Policy Issues in the Livestock Sector: Towards a Framework for Efficient and Sustainable Growth." Sector Report 8750-BR. Country Department I, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1993. "Nicaragua Agricultural Technology and Land Management Project Staff Appraisal Report." Report 11802-NI. Country Department II, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1994a. "Agricultural Extension. Lessons from Completed Projects." Report 13000. Operations Evaluation Department, Washington, D.C. Processed

———. 1994b. "Mexico Rainfed Areas Development Project." Staff Appraisal Report 12553-ME. Country Department II, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1995a. "Colombia Agricultural Technology Project Staff Appraisal Report." Report 12236-CO. Country Department II, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1995b. "Venezuela Agricultural Extension Project." Staff Appraisal Report 13591. Country Department II, Latin America and the Caribbean Region, Washington, D.C. Processed.

———. 1996a. "From Vision to Action in the Rural Sector." Agriculture and Natural Resources Department, Washington, D.C. Processed.

———. 1996b. "India Livestock Sector Review: Enhancing Growth and Development." Report 14522-IN. Country Department II, South Asia Region, Washington, D.C. Processed

Zijp, Willem. 1991. Personal communication. World Bank, Agriculture and Natural Resources Department, Washington, D.C.

Cofinanced Public Extension in Nicaragua

Gabriel Keynan • Manuel Olin • Ariel Dinar

How effective is agricultural extension? Is it worth the vast sums governments spend to provide it, mostly as a free service, to farmers worldwide? Relatively few studies exist that measure and compare the benefits of extension activities against their costs. In the absence of such data, this pilot activity concentrated instead on demand. Would demand for extension services be high if they were no longer free? The existence of solid demand would presuppose some benefits from the service. Further, might charging for the service actually improve its quality and sustainability?

The pilot program in Nicaragua described here set out to test whether a truly demand-driven extension system aimed at farmers with small- and medium-size holdings could be developed. The principal mechanism was a contribution by the farmer paid as a bonus to the extensionist; the aim was to introduce incentives for providers to improve the service through rewards linked to the quality of their work and to establish direct accountability of extensionist to client. The outcomes showed that the cofinancing concept can be successful. The article describes the design, implementation, and results in the expectation that the lessons learned may be of interest elsewhere.

Governments have spent large sums providing agricultural extension as a free service to farmers—some US$6 billion worldwide in 1988 alone, according to estimates from the Food and Agriculture Organization (1990). Is expenditure on that scale justified by commensurate benefits?

Evidence does exist that extension has increased productivity and income (Birkhaeuser, Evenson, and Feder 1991; Bindlish and Evenson 1993; Bindlish, Evenson, and Gbetibouo 1993), but it is not clear that the investigated cases are typical. Evaluation of the cost-benefit relationships has been surprisingly thin. As Birkhaeuser, Evenson, and Feder (1991, p. 643) observe: "Given that an extension organization exists in almost every country and in view of the large volumes of public funds directed to extension, there is scope for much more

World Bank Research Observer, vol. 12, no. 2 (August 199?), pp. 225–4?

empirical work on this issue." In the same vein, Purcell (1994, p. 10) notes (of World Bank–financed projects) that "quantification of the impact of extension investments in economic terms is normally not undertaken in either ex-ante or ex-post analysis."

In the absence of data, unsatisfactory results are frequently assumed: ". . . the poor performance of many public extension programs . . ." (Umali 1996, p. 2), and "the public sector extension services in which developing countries often at the behest of donors have invested large sums are achieving only limited impact . . ." (Farrington 1994, p. 1). Nonetheless, there is a common conviction that, whether universally successful or not, extension is the only means available for increasing production "given the limits on land and irrigation and likely breakthroughs in technology, future increases are apt to . . . come from technological improvements derived from identifying, developing and applying more efficient practices" (Antholt 1994, p. 4).

Another way of deducing the value of extension services would be to find out whether farmers have expressed demand for them. Pressure from interest groups in the form of petitions or strikes when the service has been discontinued would be a fair indicator of demand (Guttman 1980; Rose-Ackerman and Evenson 1985). It is instructive, for example, that no strike took place in Nicaragua in December 1995 when the government announced a substantial cut in public expenditures that would affect the extension budget, whereas the concurrent announcement of a corresponding cut in financial support to universities was met with a very violent strike.

The farmers' lack of reaction suggests limited interest, even apathy toward the service and obviously therefore calls into question its impact on farmers' economic condition. Because failure to react to removal or reduction of extension is common to most extension services in the developing world, it is likely that the services provided often leave farmers indifferent to them.

An extension service should be designed to make the outcome of its actions advantageous to both the farmer and the extension institution. Without a stake in the outcome, the extensionist will make little effort: "The amount of effort . . . depends on the value of a reward and the probability of receiving the reward. The perceived effort and probability of actually getting a reward are, in turn, also influenced by the record of actual performance" according to the Porter and Lawler model of motivation (Koontz and Weihrich 1990, p. 326).

Accountability, which Kessides (1993, p. 17) defines as "the ability of service providers to serve the interests of users and other financiers," also needs to be built into the design of the service. "Competitive markets meet this criterion of accountability in that the profit-oriented suppliers have an incentive to satisfy the demands of their customers" (Kessides 1993, p. 17). In standard extension organizations, these incentives are lacking. Data on performance are rarely col

lected, so potential reward cannot be related to any record of performance, and public institutions in any case rarely offer performance-based incentives to their employees. Consequently, little effort is likely to be expended toward satisfying clients' needs.

This article describes the methodology applied in establishing and delivering a demand-driven, accountable, and possibly sustainable extension system. A pilot activity to test the system was carried out in Nicaragua in 1995 in the context of the Agricultural Technology and Land Management Project (ATLMP), financed by the World Bank and the Swiss government. The intention of the project was to make the extensionist accountable to the client through incentives directly related to the client's satisfaction with the service. A key mechanism is the client's power to influence the quality of the service through "exit" and "voice" (Hirschman 1970)—dissatisfied farmers will cease to be clients (exit) or demand a replacement of the service provider (voice). In this design the drop in the extensionist's income that would result from exit or the threatened decline that would result from voice may be expected to induce a corrective or recuperative action. Such mechanisms do not exist in a standard extension organization.

To put the account in context, the article first reports on extension experiences in countries other than Nicaragua in which farmers share in the cost of the service and goes on to provide some background on the extension service in Nicaragua as a whole. Subsequent sections discuss the methodology used in the pilot, the results and lessons learned, and authors' conclusions.

Cofinanced Public Extension in Various Countries

Farmers now share the costs of extension services to varying degrees in several industrial and developing countries, a shift in policy intended primarily to reduce costs (see "The Public and Private Sector in Agricultural Extension" by Dina Umali-Deininger in this issue, pp. 203 to 24[illegible]).

Wilson (1991) reports on several models of cost sharing in Latin America. In Chile INDAP (the National Agricultural Development Institute) has contracted with consulting firms to provide extension services for a period limited to three to five years. In this plan farmers in higher income categories are required to pay 15 percent of the cost, financed by a credit provided by INDAP. In Mexico farmers at higher income levels are currently required to pay 15 percent of the cost of extension services, a share scheduled to increase gradually until it reaches [illegible]0 percent. In the Imbabura region of Ecuador, the extension agents purchase inputs and sharecrop with the farmers; the extentionist's share covers the interest on the loan for the purchase, the risk involved in the enterprise, and a payment for the services provided.

In Colombia responsibility for providing extension has been devolved to th municipalities, financed mainly by the central government with a share covere by the municipality (Garfield, Guadagni, and Moreau 1996). In Costa Rica pilot project, planned to evaluate trade in extension vouchers, awards vouche for extension services to farmers on the basis of the type of farm and level technology; farmers may trade vouchers based on their needs (Ameur 1994). I Argentina groups of voluntarily formed and self-supporting neighboring farn ers and ranchers organize to promote their common interests. Each group hir and pays for the services of an agricultural advisor (Garcia Tobar 1996). I China individual farmers or farmer associations contract with research institu tions, universities, and individual scientists to provide technical assistance (Amet 1994).

In Europe cost sharing is common. About three-quarters of the operatin extension budget in France is collected at the farm level through direct pay ments, contributions of agricultural organizations, and other direct and ind rect taxes on agricultural inputs and products (Ameur 1994). The U.K. exter sion agency, which remains partly government-funded, now charges for som services, originally offered free of charge, on a time-cost basis (Dancey 199 Ingram 1992, pp. 51–58). The Dutch Extension Service began a privatizatio process in 1993 in which the share of the extension budget to be funded by th farmers increased from nothing in 1993 to 60 percent in 1996; in 1998 farme are expected to cover 80 percent of the budget (Tacken 1996).

Since 1990 extension in Queensland, Australia, has been project-based, wit up to 30 percent of the budget funded by the clients. The cost-sharing formul for the cofinanced projects is set in negotiations; the government typically pay fixed costs such as salaries and equipment, while the farmers provide in-kin contributions such as the use of farm equipment, demonstration sites, and live stock (Coffey and Clark 1996).

Extension in Nicaragua

In 1994 agriculture accounted for about one-fourth of Nicaragua's gross do mestic product and employed about one-third of the labor force (Banco Cen tral 1996). The total number of farmers was estimated at some 250,000 (Centr de Investigaciones 1988). Agricultural products, mainly coffee, sugar, and beel account for as much as three-fourths of the country's total exports.

Public agricultural extension in Nicaragua dates back to at least 1942, whe the U.S. Department of Agriculture under cooperative agreements establishe two experimental stations in the country (Hernandez 1991). The extensio service underwent numerous conceptual and organizational changes over th

years, but at all times remained highly dependent on external financing. Although public extension was funded through the public budget, the government's financial contribution was always limited. In 1996, for example, almost 60 percent of the basic extension budget was financed from external sources (of this, 80 percent was from a World Bank loan and the rest from Canada, Holland, Japan, Norway, Switzerland, and the European Community). Extension activity waxed and waned according to the fluctuations in foreign aid.

The Instituto Nicaraguense de Tecnologia Agropecuaria (INTA) was created in 1993 as a semi-autonomous institution, primarily to remove it from political influences in staffing and to define it as a professional organization serving agriculture. It is divided geographically into five regions with its headquarters in Managua. The regions are all located in the west (Pacific) and center of the country where economic and agricultural activity have traditionally been concentrated.

In 1996 INTA employed some 160 extensionists throughout the country, serving some 21,500 farmers in its five regions (INTA 1996). Another 47 extensionists provided service to 5,400 farmers through a private technical assistance (PTA) program cofinanced by the government and contributions from the farmers, who are expected eventually to undertake most of the cost. In total, INTA's various programs reach about 27,000 farmers (about 11 percent of the country's total)

INTA's service is free, with the exception of the program described in this article and the one provided through five private firms in the PTA. Under the PTA program the farmers were to cover some 20 percent of the cost in the first year (1996), and their contribution was scheduled to rise to cover most of the costs of the service within five years. In 1997 no farmer was paying more than 30 percent of the cost. An estimate of the average cost of public extension provision in Nicaragua (based on the INTA budget for 1995) suggests a cost per farmer of $115 a year.

Extension is also provided to a very limited extent by other bodies such as UNICAFE, the coffee growers association; UNAG, an association of primarily small farmers, and numerous nongovernment organizations (NGOs). In December 1995 UNICAFE reported employing fifteen full-time extensionists, while APENN (national NGO) employed three. UNAG employed only two technicians in 1994, but also fielded a larger number of promoters in their "Campesino a Campesino" (farmer-to-farmer) program. (The promoters have very limited training.) NGOs generally provide extension free of charge. UNAG's program requires the farmers to offer a meal to the visiting promoter. The UNICAFE program is financed through a charge on coffee exports and is in effect a tax on production. Extension services of all types reach only a small proportion of the farmers of the country.

Designing and Planning Cofinanced Public Extension in Nicaragua

In 1994 INTA's budgetary difficulties forced the agency to rethink its relationships with its clients and the way it provided extension services. In particular, the agency thought it necessary to create accountability in the service and to institute payments by farmers that would make it possible to reward extensionists who delivered good results.

The Concept

With those objectives in view, the ATTMP sponsors—the World Bank and the Swiss government—sought to develop a demand-driven extension system. The principal shortcomings of the existing extension service to be corrected were:

- Unclear objectives: extensionists did not have a clear sense of what they were expected to accomplish.
- Poorly motivated workers and management: there were no incentives to produce results.
- No accountability to clients.
- Little consumer interest in obtaining quality service.

The design of a delivery system for demand-driven extension services was based on the following assumptions:

- Extension is an economic input.
- Extension generates new income.
- Farmers, even if poor, will be willing to pay for an input whose expected value is greater than its cost.

The objective of extension in this concept is to help farmers who receive the service to increase incomes derived from agricultural activities by increasing yields, reducing losses in the field and after harvest, reducing costs, improving exploitation of available resources, and designing a better mix of products.

The principal mechanism proposed by the sponsors to achieve the objectives and correct the shortcomings was to charge for the service, on the rationale that payment serves the following purposes:

- Extension staff begin to regard farmers as clients to whom they are accountable.
- When their remuneration is linked to economic results, extensionists have a stake in the farmer's success.
- People are committed to a service they pay for; they are careless of a service they get free.

Promoting the Concept of Cofinancing

When payment for public extension service was first broached in February 1994, the Nicaraguan authorities and the leadership of the major agricultural associations rejected it as unworkable, unfair, and contrary to tradition. Implementing a payment system over such objections would clearly have been impossible, so a discourse with all the various interest groups was initiated to persuade them to withdraw their opposition. Project sponsors approached leading officials in the Ministry of Agriculture and Livestock, the board of directors of INTA, and the leaders of the major agricultural associations individually, in each case presenting the rationale for payment, accompanied by many examples related to the work of the institution being visited. Free discussion generally led to grudging acceptance and finally to agreement. Ultimately, this campaign was so successful that cofinanced public extension is now accepted by all sectors, including the associations of small farmers.

The next step was to determine the reactions of the principal players—the extensionists and the farmers—and to identify, understand, and deal with any reservations and qualms that they might have. Formal and informal discussions with extensionists and other staff in INTA extension districts revealed that extensionists lacked self-confidence and conviction about the effectiveness of extension and its impact on farmers' income. They doubted that any farmer would be willing to pay for it.

To help build confidence and to deepen their understanding of extension objectives and performance, extensionists were invited to participate in group discussions with INTA management and sponsors in their regions. Using actual cases suggested by the extensionists themselves, they were guided on how to identify benefits from their service and to calculate the value of the advice they had given to farmers. Directed questioning helped them think through a process and assign a value to their work. These discussions succeeded in increasing their confidence and reducing opposition to the concept.

At the same time, meetings were held with small groups of farmers throughout the country. Following the same format and process as the meetings with the extensionists, these discussions helped farmers identify some benefits of the extension that they had received and place an economic value on the service. At the conclusion of these meetings, they were asked if they would be willing to pay for a service that increased their income. Although none evinced enthusiasm, in all but one case the farmers agreed that they were. These discussions served several purposes simultaneously: they helped farmers to place an economic value on extension; they showed extensionists (who attended the meetings) that farmers would not reject the idea of payment; and they sent a message to farmers that payment for the service was being considered. These discus-

sions, which set the stage for implementation, were carried out intermitten over several months before the concept was tested in the field.

Concurrently, INTA management, prodded by the sponsors to reflect on t probable economic situation in which the institution would no longer ha access to external loans, generally acknowledged that the government could n maintain funding at the current rate, and that payment by farmers for exte sion services might help to fill the gap.

The Field Test

The field test was carried out in 1995 in the *postrera* season (the second Nicaragua's two agricultural seasons: *primera*, which begins about May; a *postrera*, which begins in August). The purpose of the field test was to find o whether farmers would in practice pay for agricultural extension services—th had accepted the concept in theory in the previous round, but no one was ask to "put your money where your mouth is." The objects of the exercise were test farmers' reaction, to show that extension has economic value, to gain farm confidence by delivering on promises, and to establish the principle that t service is not free. It was decided to test the principles on four farmer groups two regions.

In preparation for the test, a workshop was held with staff from the tv regions, Esteli and Matagalpa. After the concept was presented, virtually participants declared that it could never work. Although all had participated the earlier discussions, the threat of actual implementation on their own tu again raised opposition. By the end of the day's discussion, however, there w actual enthusiasm for the idea and a genuine willingness to test it. The turni point was the clarification that the extensionists would retain all the mon received from the farmers.

Because farmers would join the program voluntarily, the concept would ha to be "sold" to them. They would have to be convinced that it was to the benefit to pay for the service. The guiding principles for the test were volunta participation, no threats that existing services would be discontinued, negoti tion on services to be given and received, and negotiation on price and terms payment. Because only these four farmer groups would be paying, the packa offered for payment had to be superior to the package given free of charge to other farmers.

The workshop identified potentially interested groups, a short outline what they would be offered, and how much they would be asked to pay (tab 1). All agreed that the farmers most open to such a program would be tho

Table 1. ***Field Test, 1995: Payment Scale for Major Crops***

Crop	*Payment scale (córdobas per manzana per month)*[a]	*Duration of season (months)*
Tomatoes	10	3
Beans	6	3
Maize	4	4

a. At the time of the field test, US$1.00 = 7 cordobas, one manzana = 0.7 hectares.
Source: Data collected by INTA staff and processed by the authors.

who raised a cash crop, who had access to some credit, and whose yields could readily be increased by better farm management.

The prices set had to meet the following criteria: the sum had to be one to which a farmer would not be indifferent; and the total income expected (the sum of all payments) had to be a sum that would constitute an incentive to an extensionist. The prices were unrelated to either costs or benefits. Rather they were based on the sum paid monthly by low-income families for electricity. This was felt to be an indicator both of their ability to pay and of how much they would be willing to pay for a service promising some benefit.

Each region was permitted a great deal of latitude on the services offered and the rate charged. Cost recovery was not an objective either of the test or of the concept in general. Each region was free to propose any program within the scope that INTA thought to be attractive to the farmers, provided that the services offered did not go beyond those that extension traditionally provides. The essence of the package offered was greater frequency of visits of extensionists to farmer groups and included the right to a few visits to the plots of individual farmers. Extensionists would not offer advice on what crops to grow unless specifically requested, and not before gaining the confidence of the farmers by helping them to achieve better results than they traditionally obtained.

The selling process consisted of meetings with the farmer groups originally targeted, as well as others added later. The farmers accepted the idea of payment with relative ease. They generally claimed to have no access to credit, however, and maintained that without credit they would be unable to pay for the service, it soon became clear that they would not participate without being given access to credit. In Matagalpa the groups approached were also targeted by a local NGO dedicated to the export of nontraditional crops. This NGO offered intensive extension along with credit, land preparation, and marketing. The farmers rejected INTA's offer as inferior. No alternative groups were located in Matagalpa, probably because the INTA team felt weak in comparison with the "competition" and was demoralized by the failure of its first attempt.

In Esteli the farmers also conditioned their participation in the scheme on access to credit. INTA's regional director took the initiative of directing the funds for seed production, which he had available from another NGO, to groups of farmers that would agree to participate in the test. Two groups accepted the offer and entered into an agreement with INTA.

The agreement was basically between the farmers and the extensionist; INTA backed it. The farmers wanted a weekly visit to the group (rather than the customary biweekly visit) and asked for visits to individual fields on demand. Because the groups were centrally located and readily accessible, this request was easy to accommodate. INTA required the extensionist to devote some personal time to this service, including visits on Saturdays and Sundays.

Payment would be on the basis of the crop rates suggested above (see table 1), but only half the sum collected would go to the extensionist. The farmer groups had decided to use this opportunity to build up their own capital by retaining the other half in a fund. They also requested that payment be based only on some of their crops; they did not feel the need for advice on the farm as a unit. These changes were not desirable but were approved to get a field test of the system up and running.

A meeting to evaluate the experience was held by INTA managers and the sponsors at the end of the season to hear reports by the farmers and the extensionist. The farmers reported that the extensionist had visited even more than once a week. They also readily acknowledged that the test had been a success and that they were satisfied with the results. The most notable improvements were in tomatoes. They reported harvesting about 800 boxes of tomatoes per manzana instead of the 600 boxes they traditionally harvested, a difference they ascribed to extension. The extensionist, however, was not content with the return for his efforts. He had devoted a great deal of his own time to the test and had assumed that he would earn a higher sum than he ultimately did—the equivalent of about one week's salary for the three-month season (table 2). He acknowledged, however, that the field test was a preparatory phase and

Table 2. *Amounts Paid to Extensionist*

Crop	Area (manzanas)	Rate (córdobas/month)	Duration of season (months)	Total paid (córdobas)	Paid to extensionist (córdobas)
Tomatoes	12	10	3	360	180
Beans	18	6	3	324	162
Maize	6	4	4	96	48
Total	n.a.	n.a	n.a	780	390

n.a. Not applicable.
Source: Data collected by INTA staff and processed by the authors

showed promise for the future. He was also proud of the reputation he had gained among his clients and of his status among his peers.

The farmers were asked how they had arrived at the decision to pay the sum of 5 córdobas per manzana per month and whether they did not feel that a higher sum was justified. They replied that they were too poor to spend more on the service. Upon further questioning, they reported that the average price received was about 15 córdobas per box of tomatoes, giving a gross additional income of about 3,000 córdobas per manzana. They calculated their additional costs (mainly for boxes and transportation) at some 1,800 córdobas, leaving them with a net income of about 1,200 córdobas per manzana. They had earned this sum in return for an expenditure of about 15 córdobas per manzana (three months at 5 córdobas). When presented with this calculation, they were surprised at the extent of the economic benefit and agreed that it was far in excess of what they had paid.

In proving that the approach was feasible, albeit on a very small scale, the positive results of the test gave a strong impetus to the program. When it was announced that the trial would be expanded and a pilot carried out during the following season on a larger scale, extensionists throughout the country were eager to participate.

The Pilot

The pilot program was launched in the *primera* of 1996, after completion of the test, building on the lessons learned to apply the concept on a larger scale.

Strategy

Before broadening the scope, a strategy had to be developed for introducing cofinancing into forward planning for public extension as a whole. The plan defined three classes of service: free (traditional); participative (cofinanced, carried out by INTA's staff); and private (cofinanced, carried out by private consulting firms). The essence of the strategy was that future extension would end up as one of the two cofinanced modes after a process of transfer from free to cofinanced extension. Free extension was to be limited to a maximum period of two to three years, after which the farmer would either transfer to cofinanced extension or would cease to receive the service.

Free extension was to be viewed as a "free sample" to interest new clients in the service and would at any time in the future be only a small percentage of the total clientele. For current clients, free extension would also be phased out and the service either transferred to one of the cofinanced modes or discontinued,

but a decision as to the date was postponed until more experience with the concept could be gained.

Tools

Extensionists were trained to use tools developed for use in the pilot, principally the preparation of contracts, forward planning of both farm and extension activities, measurement of results, and monitoring of progress.

The *contract* formalizes the arrangement between a group of farmers and the extensionists: what services are to be provided at what price. The contract establishes the seller-buyer relationship between the extensionist and the farmer It contains an outline of the services that a group of farmers might want and the price thought acceptable. It is meant to be revised after negotiation.

Rapid analysis of the farmers' resources and constraints, to be based on information about the fundamental features of the farm, gathered using a standard data collection form, is a tool for extensionists to use in planning and advising farmers.

Farm planning in this context does not imply a farm plan in the normal sense; rather, it is a means of setting objectives in order to create a "promise" on the part of the extensionist and an "expectation" on the part of the farmers, as well as a benchmark for measuring the success of the service. The form used places the "traditional" yield and the "promised" yield side by side.

Plans for the extensionist's activities must be relevant to the farm plans and not general to the extension service. On this basis the extensionist selects and plans the subjects to instruct in, the types of field demonstrations relevant to the farmer groups, and where and when to carry them out.

The farmers themselves must *measure results* because the extensionist cannot be present at all harvests and all times of production. Measurement by the farmer includes both products sold and those used in home consumption and must be continuous or intermittent, depending on the crop. Extensionists are advised to check regularly to see whether the forms are being filled in and whether the farmer has encountered any difficulties in this activity. After all harvesting is concluded, the farmer and extensionist should jointly estimate the value of increased yields, calculate and deduct any additional costs associated with the recommendations, and finally calculate the outcome.

An intense *program of support and monitoring* was prepared, to be carried out by the pilot promotion team and a person assigned to provide support in each region, selected because of his or her dynamism and enthusiasm for the program. This support person would maintain regular contacts with the extensionist and the groups. The promotion team would visit each region once a month, and the support persons from all regions would congregate once a month in a

different region, coinciding with one of the promotion team's visits. The meetings would include sessions with the extensionists and with the farmer groups. In addition a workshop was programmed for the end of the season.

Results

Eighteen groups signed up to participate in the pilot by early 1996. Of these, seventeen remained and received extension services throughout the season (the eighteenth dropped out because of early flooding of its fields). Some individual farmers who had misinterpreted the payment for extension as meaning that it was a credit program, dropped out of the groups. Ultimately, 280 farmers were served. Several changes and adjustments occurred during the pilot. Many of the agreements changed after signing, partly because early losses of crops obliged the farmers to revise their original farm plans late in the season. Although such revisions may be common, their design and the provision of advice for the new plans constituted an important part of extension work. Another element of the original design, the rapid analysis form, was dropped because it proved so time consuming to complete.

A breakdown of the area grown by different crops and by farmer groups is shown in table 3; the results, reported in terms of yields and net benefits to farmers, are presented in tables 4 and 5; and sums payable to extensionists and the actual payments (by June 1996) appear in table 6.

Table 3 shows that more than 50 percent of the area was devoted to beans, a crop grown by ten of the seventeen farmer groups. The three major crops—beans, corn, and rice—covered 85 percent of the total area of the pilot.

Table 4 shows the data on basic crops reported by twelve of the seventeen groups. Four others specialized in cattle raising or, less commonly, in vegetables.

Table 3. *Crops Contracted*

Crop	*Percentage of area*	*Number of groups with this crop*
Beans	54	10
Corn	22	8
Rice	9	2
Tomatoes	6	6
Sesame	4	2
Sorghum	3	4
Other crops	2	4
Livestock	n.a.	4

n.a. Not applicable.
Source: Data collected by INTA staff and processed by the authors.

Table 4. *Net Incremental Income by Regions and Crops*

Region/group	Number of farmers	Crop	Area (manzanas)	Yield (quintal/mz) Traditional	Yield (quintal/mz) Actual	Farm incremental yield (quintal)	Price per unit of yield (córdobas)	Farm: Gross incremental income (córdobas)	Farm: Incremental production costs (córdobas)	Farm: Net incremental income (córdobas)
A-1/Quetzalgue	23	Corn	6	20.0	10.0	−60.0	55	−3,300	−690	−2,610
A-2/Masaya	27	Beans	27	9.0	10.0	27.0	220	5,940	4,430	1,510
A-2/Carazo	16	Beans	15	12.0	9.8	−33.0	500	−16,500	1,530	−18,030
A-2/Ticuantepe	16	Beans	18	9.0	13.9	88.2	250	22,050	3,714	18,336
		Corn	1	20.0	20.0	—	100	0	205	−205
A-2/Cardenas	19	Rice	26	64.0	68.9	127.4	140	17,836	11,401	6,435
B-3/Jalapa	24	Beans	24	8.0	6.0	−48.0	180	−8,640	2,760	−11,400
B-3/Ocotal	25	Beans	4	6.5	10.0	14.0	150	2,100	137	1,963
		Corn	22	26.0	41.1	332.2	55	18,271	3,713	14,558
B-3/Condega	22	Beans	15	8.0	3.0	−75.0	150	−11,250	3,375	−14,625
B-5/Barro	31	Beans	21	12.0	8.6	−71.4	200	−14,280	5,607	−19,887
		Corn	19	15.0	14.2	−15.2	60	−912	2,701	−3,613
B-5/Cebadilla	15	Beans	17	8.0	5.3	−45.9	200	−9,180	3,288	−12,468
		Corn	12	12.0	7.1	−58.8	40	−2,352	−4,106	−6,458
C-6/Boaco Viejo	16	Beans	12	15.0	19.1	49.2	200	9,840	143	9,697
		Corn	10	15.0	18.5	35.0	40	1,400	120	1,280
C-6/La Libertad	9	Corn	2	20.0	18.7	−2.6	80	−208	400	−608

Note Corn, beans and rice only

Source Data collected by INTA staff and processed by the authors

One group in Region C-6 did not provide information on its results. Of the sixteen groups reporting, eight recorded net gains and eight net losses. Three of the groups that had losses, reported increased income in one or more of the crops. Table 5 summarizes the results reported by all groups by region.

The quality of the data received from the fields was questionable. Some data on yields and income were inconsistent and had unexplained variations among farmers. For example, in some cases the yields and the net benefits with extension were reported to be lower than without this service. This would be a very unfortunate fact indeed, if true. But, because almost all farmers reported their satisfaction with the service and signed contracts for the following season, it could not have been true. The more likely conclusion is that farmers were not aboveboard, overestimating traditional yields and underreporting actual yields; they probably also inflated their traditional costs of production. The potential for such misreporting revealed a weakness in the reporting system as well as a professional weakness among the extensionists, who were themselves unfamiliar with traditional yields and unable to estimate visually the expected yields by observing the fields during the season.

Some data on drop in yields, however, were correct. On some farms a variety of beans was promoted that was not suitable to the prevailing conditions. Also, in some areas flooding and late planting because of heavy rains clearly affected the yields.

The farmers paid more than 60 percent of their fees within a reasonable time (June 1996; see table 6), indicating that they were willing and able to pay. Recovery might have been higher, but many extensionists were reluctant to collect, feeling uncomfortable with this activity. Most of the remaining farmers ultimately paid their outstanding balances, because all were informed that ful-

Table 5 *Net Incremental Income by Farmer Groups*

	Number of farmers	Area of crop (manzanas)	Number of livestock	Value of difference (córdobas)[a]	Incremental production costs (córdobas)[b]	Net incremental income (córdobas)
[illegible]	55	[illegible]	95	13,710	170	−13,540
[illegible]	78	88		28,808	21,389	7,419
[illegible]	75	[illegible]		22,457	14,789	7,668
[illegible]	46			38,488	16,062	54,550
[illegible]	26	35	362	114,288	14,836	99,452
Total	280	298	457	113,355	66,906	46,449

a. Difference between previous season yields (*primera* 1995) and *postrera*'s 1995 yields, multiplied by [illegible] price.

b. Additional production costs in the *postrera* 1995 compared with the situation before the test (i.e., [illegible] 1995).

Source: Data collected by INTA staff and processed by the authors.

Table 6. *Payments for Extension Services, Postrera 1995*

Region	*Total agreed (córdobas)*	*Total paid by June 1996 (córdobas)*	*Percentage paid*	*Range of percentage paid by group*
A-1	3,258	1,457	44.7	20.7–61.5
A-2	2,056	1,594	77.5	47.8–100.0
B-3	3,970	2,820	71.0	0.0–100.0
B-5	1,410	1,410	100.0	—
C-6	1,061	158	14.9	0.0–27.0
Total	11,755	7,439	63.3	-

— Not available.

Source: Data collected by INTA staff and processed by the authors.

filling their payment agreements was a condition for receiving the service in the following season.

Lessons Learned

Several preliminary lessons drawn from the pilot were applied in the subsequent season and may guide other countries interested in similar application of cofinanced extension

The System Works

Farmers agreed to pay for the service, and most of them paid as agreed. They uniformly voiced their satisfaction in discussion with INTA staff and the pilot sponsors and voted with their feet by signing new contracts for the following season. The most common comment was that the extensionist was much more attentive and responsive to their needs than in the past. This had indeed been the primary objective of the concept.

The additional income extensionists earned in the course of the pilot instilled both a desire to seek out more clients and to serve existing clients better. A notable expression of this spirit was pressure by the extensionists to decrease the number of days devoted to training (which was reduced by management to a maximum of two days per month), whereas previously they had pleaded to be included in all training and workshops offered regardless of how this would affect their visits to the farmers.

The principal indicator of the success of the method is that all seventeen original groups have continued and by the *postrera* of 1996 (one year after the

testing of the pilot was initiated), the number had increased to 135 groups with 1,737 farmers.

The Process of Persuasion Is Critical

The sponsors proposed a concept that held high promise but that was also contrary to tradition and unpopular politically. Identifying the sources and reasons for the opposition, understanding the nature of the objections, resolving problems, and making allies of the opponents, all before any action was taken, were crucial steps to success. Actors at all levels, including farmers, participated in discussions and in developing the practical aspects and tools of the system. Many of their ideas were accepted and integrated, even when serious doubts existed about their viability or desirability, as long as they were not contrary to the basic principles. As long as the idea did no damage and could be corrected later, the support of the participants was considered to be paramount. Care was taken to ensure that the staff never felt that the concept was being imposed on them but rather that they were part of it.

It was also essential to avoid pushing too hard or too fast. Ideas are absorbed gradually, and even when understood intellectually, they are not truly assimilated until they become reality. Quantitative goals projected more than one season ahead appeared too outlandish and only indicated to INTA's staff and the farmers that the speaker did not have his feet on the ground. In the event, later goals were easily achieved and even surpassed.

The deep commitment and conviction of the sponsors and their perseverance was another element critical in the success. A high percentage of the time devoted to this pilot went into jawboning. Sponsors resisted the temptation to rush headlong into field testing, delaying implementation until the entire stage was properly set.

Start with a Comprehensive Strategy

From the outset the philosophy underlying the pilot was clearly articulated and concentrated on a small number of principles. Although the pilot was small, the sponsors and INTA management regarded it as a test of the methodology of the future, and consequently of interest to all—never as an activity affecting only a few staffers and a few clients. INTA management agreed to a strategy for full implementation of the concept, even while the pilot was little more than an idea. Thus the pilot was always part of the big picture and staff members knew where they were heading.

Staff at all levels participated in discussions on the whys and wherefores of the approach. No one was asked to do anything without first receiving explana-

tions of the logic behind it. Furthermore, detailed tools for all the major sta[g
were developed and training given in them, even though the pilot was sm[a
Many more staffers were trained than would actually participate in the pilo[t

Key Issues

The pilot proved that the concept was feasible and accomplished its purpos[e
But the pilot also highlighted problems and issues that needed to be dealt wi[th
in order to refine the concept and make it more generally applicable.

Ability and willingness to take initiative and to adapt. Although i[n-
formed that the typical contract was a model and had to be adapted to t[he
specific agreement negotiated between the extensionist and the farmer grou[p,
virtually no extensionist modified a contract. More accustomed to followi[ng
instructions than to adapting to the specific needs of their clients, they retain[ed
the original form down to ". . . visiting every other Thursday . . . ," not ev[en
changing the day. This is a big problem, with implications far beyond its mani-
festation in the issue of contract modification, and requires both better st[aff
selection and training to make extensionists capable of showing more initiativ[e.

Pricing policy, price of service, and timing of payment. The matter [of
pricing was approached pragmatically and allowed for many variations. T[he
only firm requirement was that some payment be made for the service. The[re
was never any intention to charge by cost of service. The sponsors had set out [to
test whether farmers would pay for extension, and whether this would mak[e
extensionists responsive to their clients, not whether the full cost could be r[e-
covered from the clients. It was felt that realistic prices could be defined on[ly
after sufficient experience had been gained, and any price set at the outset wou[ld
be no more than a "bargaining price." Payment was set by crop and area duri[ng
the course of the growing season, because the growing season sets the cycle [of
farm expenditure and earning and the policy would therefore be explicable an[d
acceptable to farmers; farmers never questioned this policy.

Payment was divided into two categories: by crop and time (payable through-
out the season); and as a percentage of increased yield (payable at the end of th[e
season). The payment as a percentage of incremental yields is problematica[l
because the greater the increase in yield, the greater the payment. The farme[r
then has good reason to withhold information on higher yields and, becau[se
the extensionist cannot be present at every harvest, a potential for misreportin[g
is introduced. The sponsors agreed to include this form of payment despit[e
reservations; in the event, this was probably partly responsible for th[e
misreporting of yields.

Pricing the products also presented some problems. Many products are sold over a period of time during which prices vary; the farmer does not necessarily keep accurate track of the prices received nor of the quantities sold at each time—and is unlikely in any case to share this information with the extensionist. And how to deal with products stored and sold later? Those who sell immediately upon harvest receive a lower price than those who have invested in storage facilities and are able to hold on until the price rises. Should the extensionist share in the increased profit?

Having decided what to charge for the service, the question remained when the payment should be made. To offset the risk of default, the sponsors wanted payment to be made throughout the season, not only at the end. Initially there was some resistance to this, but ultimately more than half of the sums were paid during the season. It is expected that as both farmers and extensionists become more accustomed to the system, the bulk of the payments will be made during the season (with the exception, obviously, of the portion based on yield increases).

MEASUREMENT AND REPORTING. The method of measurement applied in this pilot was oversimplified and inaccurate, but it did accomplish its main objectives. It made the farmers aware of the need to measure the value of the service, put pressure on the extensionists to perform, and served as an important exercise in analyzing information and trying to learn from the results. The method deliberately ignored the fact that insofar as extension delivers a previously unfamiliar technology that increases yields and income, the flow of benefits continues for many years—in other words, that the actual value gained by the farmer is far in excess of the net benefit gained in one season. Only those benefits that could be measured easily and only improvements in the "cofinanced" crops themselves were measured. More complex, difficult-to-measure phenomena and benefits gained as a by-product in other activities were excluded. There is then a strong downward bias of the value. Farmers are reminded that the benefit is "at least such and such a sum."

One weakness identified in the reporting system was that management was undemanding about the quality of the data, primarily because of inexperience in actually using the data in its decisionmaking. Extensionists were aware of this laxity and therefore tended to supply figures with little basis in fact. Strong action was taken to correct this. Cavalier reporting will probably be a problem in other places as well and should be dealt with early.

PROFESSIONAL QUALIFICATIONS. In the course of the pilot, the promotion team focused only on the method of delivery and on its management and organization. It was assumed that the extensionists were already equipped with know-

ledge on a range of agricultural technologies, farm problems, and their solutions, and had the necessary ability and experience to diagnose problems adequately and to suggest remedies. These assumptions were found to be only partially sound. Farmers expect the extensionists to be at a very high professional level. Extensionists cannot simply be conveyors of messages to farmers. They must be able to provide advice on more complex farm management issues. Farmers' expectations are important both in evaluating the resources necessary to provide such a service and in calculating its cost. The service cannot work with poorly trained technicians.

As well as training in farm management and basic economic analysis, training for extensionists in elementary marketing skills to sell the service to potential clients was also found necessary. When establishing a new system extensionists should be selected according to their training, experience, and proven abilities to analyze and diagnose. Well-trained people are likely to be available in any given country.

Participation of women. Women constituted some 21 percent of the clients in this pilot. (No special effort was made to target women.) Although INTA does not collect data on the number of women it serves under its regular program, indirect indicators show that the pilot served about the same number of women as the regular program. The potential impact on women in a demand driven extension system is a subject that merits further study

Distribution of income among supporting staff. The scope of the pilot was too small to apply all elements of a fully operational system. Distributing the small sums collected among the staff would have nullified the incentive effect; consequently, the entire payment remained with the extensionist. The long-range plan, however, is to distribute the income widely among the staff members who contribute to the effort, as more clients are added and income increases. The extensionists depend on subject-matter specialists for problem solution, on administrative staff for transport, and on others to enable them to provide a satisfactory service to their clients. A formula has been developed for distribution as income increased, to be applied in subsequent stages.

Conclusions

The sponsors of the pilot set out to test whether a truly demand driven extension system aimed at farmers with small- and medium-size holdings could be developed. Central to the design was the principle of accountability of extensionist to client: the farmer would be charged for services received, and

the extensionist would benefit in direct relation to this payment. The incentive system thus introduced proved effective: by the end of the season, 60 percent of farmers had paid for the service, and all of the rest had agreed to pay their debts as a condition for receiving the service in the following season. The principal indicators of success are the repeat contracts with all participating groups, and the continuing expansion of the service (from 17 groups and 280 farmers in the 1995 pilot to 135 groups with 1,737 farmers in 1996). Critical to the project's viability was the period of intense activity before field testing dedicated to persuading the various actors of the viability of the concept, identifying and analyzing opposition, and making necessary corrections and adaptations. Without this careful stage-setting, a cofinanced system would have had little chance of getting off the drawing board.

The success—albeit on a small scale—of the pilot shows promise that the concept could contribute to the sustainability of extension activities in general. Ideally, agricultural extension services should continue to be offered as long as there is demand for them, even after World Bank–financed project funding is terminated. Given the notorious unreliability of government funding, the only sound basis for sustainable operations is demand, in the form of willingness to pay, and the returns for supplying a service, in the form of payment. The test did not set out to discover whether a demand for extension exists at a market price, but the results do indicate that farmers will pay something for the service. In most European countries, farmers pay about half the cost of extension, with government covering the other half (Ameur 1994). Clearly the prospects for sustainability are much improved if farmers are willing to pay a 50 percent [illegible].

In sum, the results of the pilot operation in Nicaragua suggest that cofinanced agricultural extension could contribute to the quality and sustainability of the [illegible] elsewhere. If the lessons learned are absorbed, the process described in this article could be followed to apply the concept successfully in a wide variety of situations in developing countries to the benefit of the farmers served by the extension service.

Notes

[illegible] Keynan is a consultant with the Sector Leadership Group of the Latin America and [illegible] Region of the World Bank. Manuel Olin is a consultant to, and Ariel Dinar is a [illegible] economist with the Agriculture and Natural Resources Department at the Bank. Funding [illegible] study was provided by the Swiss [illegible] Fund for Special Studies, in the context of [illegible] the Agricultural Technology and Land Management Project in Nicaragua. The [illegible] benefited from many comments and suggestions from Charles Ameur, Charles Antholt, [illegible] Feder, Miguel Gomez, Dennis Purcell, Andrew Spurling, John Stemp, and Willem [illegible]. Assistance in the field was provided by [illegible] Ricardo Hernandez.

References

The word "processed" describes informally reproduced works that may not be comm available through library systems.

Ameur, Charles. 1994. *Agricultural Extension: A Step Beyond the Next Step.* World Bank T nical Paper 247. Washington, D.C.

Antholt, Charles H. 1994. *Getting Ready for the Twenty-First Century: Technical Chang Institutional Modernization in Agriculture.* World Bank Technical Paper 217. Washin D.C.

Banco Central de Nicaragua. 1996. *Indicadores Economicos.* Managua, Nicaragua.

Bindlish, Vishva, and Robert E. Evenson. 1993. *Evaluation of the Performance of T&V E sion in Kenya.* World Bank Technical Paper 208. Washington, D.C.

Bindlish, Vishva, Robert E. Evenson, and Mathurin Gbetibouo. 1993. *Evaluation of th formance of T&V Extension in Burkina Faso* World Bank Technical Paper 226 Wasl ton, D.C.

Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder 1991 "The Economic Imp. Agricultural Extension: A Review." *Economic Development and Cultural Ch* 39(April):607–50.

Centro de Investigaciones Economicas de la Reforma Agraria. 1988. Gobierno de Nicar Managua.

Coffey, Shaun, and Richard Clark. 1996. "Reengineering an Extension Service " Paper sented at the Extension Workshop on Alternative Mechanisms for Funding and Deliv Extension. World Bank, Agriculture and Natural Resources Department. Washin D.C. Processed.

Dancey, R. J. 1993. "The Evolution of Agricultural Extension in England and Wales. I dential Address." *Journal of Agricultural Economics* 44(September):375 93

Farrington, John. 1994. "Public Sector Agricultural Extension Is There Life after Struc Adjustment?" ODI. *Natural Resources Perspectives* No. 2 (November)

Food and Agriculture Organization 1990 *Report of the Global Consultation on Exte* Rome.

Garcia Tobar, Julio. 1996. "CREA-AACREA Associative Private Agricultural Extension in gentina." Paper presented at the Extension Workshop on Alternative Mechanisms for F ing and Delivering Extension. World Bank, Agriculture and Natural Resources De ment, Washington, D.C. Processed.

Garfield, Elsie, M. Guadagni, and Daniel Moreau 1996 "Decentralization of Agricul Extension Services in Colombia." Paper presented at the Extension Workshop on Alte tive Mechanisms for Funding and Delivering Extension World Bank Agricultur Natural Resources Department, Washington, D.C. Processed

Guttman, J. M. 1980. "Villages as Interest Groups: The Demand for Agricultural Exte Services in India." *Kyklos* 33:122-141.

Hernandez, Julio Ricardo. 1991. *Economia y Cambio Tecnologico en la Agricultura Nicaragu* Managua, Nicaragua: INCAE (Central American Institute for Enterprise Administratio

Hirschman, Albert O. 1970. *Exit, Voice and Loyalty* Cambridge, Mass Harvard Unive Press.

Ingram, Paul. 1992. "The United Kingdom Experience in the Privatization of Extension Jock R. Anderson and Cornelis de Haan, eds., *Public and Private Roles in Agricul*

Development: Proceedings of the Twelfth Agricultural Sector Symposium. Washington, D.C.: World Bank.

INTA (Instituto Nicaraguense de Technologia Agropecuaria). 1996. *Plan Operativo Anual.* Managua.

Kessides, Christine. 1993. *Institutional Options for the Provision of Infrastructure.* World Bank Discussion Paper 212. Washington, D.C.

Koontz, Harold, and H. Weihrich. 1990. *Essentials of Management.* 5th ed. New York: McGraw Hill.

Purcell, Dennis. 1994. "Agricultural Extension: Lessons from Completed Projects." World Bank, Operations Evaluation Department, Washington, D.C. April. Processed.

Rose-Ackerman, Susan, and Robert E. Evenson. 1985. "The Political Economy of Agricultural Research and Extension: Grants, Votes, and Reapportionments." *American Journal of Agricultural Economics* 67:1-14.

[illegible]acken, Willem. 1996. "The Dutch Extension Service: The Change from a Publicly Funded Extension Service to a Privatized, Client-Oriented Organization." Paper presented at the Extension Workshop on Alternative Mechanisms for Funding and Delivering Extension. World Bank, Agriculture and Natural Resources Department, Washington, D.C. Processed.

Umali, Dina. 1996. "New Approaches to an Old Problem: The Public and Private Sector in Extension." Paper presented at the Extension Workshop on Alternative Mechanisms for Funding and Delivering Extension. World Bank, Agriculture and Natural Resources Department, Washington, D.C. Processed.

Wilson, Mark. 1991. "Reducing the Costs of Public Extension Services: Initiatives in Latin America." In William M. Rivera and Daniel J. Gustafson, eds. *Agricultural Extension Worldwide: Institutional Evolution and Forces for Change.* Amsterdam: Elsevier.

Reconsidering Agricultural Extension

Robert Picciotto • Jock R. Anderson

The institutional design of agricultural extension programs in developing countries has always been subject to heated debate. But whereas previous disagreements centered on the relative efficacy of different organizational approaches to public service delivery, the questions currently raised by agricultural policymakers and extension practitioners go far deeper. They reflect a new conception of the role of the state in the rural economy (Pasour 1990; Timmer 1991; Meerman 1997), the continuing revolution in communications technologies (Zijp 1994), and the growing influence of the "new institutional economics" in development thinking (Hoff, Braverman, and Stiglitz 1993, Klitgaard 1995).

Rural Development Antecedents

Fifty years ago agricultural extension organizations in developing countries mirrored the administrative traditions of the former colonial powers (Axinn and Thorat 1972). Like other agricultural support services, extension services were geared to producing and marketing export commodities. Accordingly, crop oriented extension programs were common. The scope of extension programs expanded in the 1950s as the newly independent states of Asia and Africa sought to increase food production and to spread the benefits of improved farming techniques more widely. As extension organizations sought to expand their reach both to more farms and to more types of farming, their production orientation weakened (Baxter, Slade, and Howell 1989; Macklin 1992; Schwartz and Kampen 1992; Anderson and Hoff 1993).

The economic strategies of these pioneering years relied on heavy state intervention, import substitution, and rapid industrialization. In this context the urban bias of development policies, the adverse terms of trade faced by farmers, and the limited stock of improved technologies available for dissemination all hindered the productivity of rural development programs. In turn, extension

programs often relied on the proposition that farming productivity was held back not so much by technological and economic constraints as by farmer apathy, inadequate social arrangements, and lack of local leadership. Often, extension agents came to be viewed as the foot soldiers of "nation-building" campaigns aimed at multiple economic and social objectives.

In the 1950s and the early 1960s, the agricultural extension service tended to be subordinated to multipurpose rural development programs. Extension agents carried out a variety of functions, ranging from credit delivery and input distribution to sundry coordination duties. And because extension agents were among the few government officials available at the village level, they were often asked to undertake clerical, statistical, or even political chores. Typically, the service had only weak connections to agricultural research

Looking back, the rural development movement was the victim of a poor enabling environment for agricultural development. Eventually, it fell into disfavor as lack of profitable technical packages and an overly broad agenda led to a thin spread of resources, excessive administrative costs, and slow agricultural production growth. Still, considering the constraints then prevalent, integrated rural development achieved notable success by nurturing local leadership, enhancing the influence of rural constituencies, and creating a physical and administrative infrastructure at the local level that proved invaluable for future production-oriented programs (OED 1988, Donaldson 1991)

The Advent of the Training and Visit System

In the late 1960s and early 1970s technology diffusion became the focus of agricultural extension. The economic rationale for the shift was powerful new high-yielding, fertilizer-responsive crop varieties were available for dissemination, and food shortages forced output prices high enough to make the use of the new technologies profitable (Lipton with Longhurst 1989) These conditions in turn created a favorable situation for the adoption of a major organizational innovation—the training and visit, or T&V, system (Benor, Harrison and Baxter 1984).

Under T&V agricultural extension was expected to act as a transmission belt between agricultural research centers and millions of small farmers The reorientation of the agricultural extension system from a desk-bound bureaucracy to a field-based, professionally motivated cadre of agents, closely connected to research and geared to the systematic promotion of improved cultural practices through a strict calendar of daily visits and weekly training sessions, amounted to a major reform of agricultural services.

Yet the degree to which the remarkable food production gains of the green revolution can be attributed to any particular institutional mechanism, such as

T&V, has long been disputed (see Raman, Balaguru, and Manikandan 1988; Lipton with Longhurst 1989; Antholt 1991; Feder and Umali 1993; Foster and Rosenzweig 1996; and Evenson, Pray, and Rosegrant forthcoming). In any event T&V has often cohabited with other means of diffusing new technologies. Even in India, the cradle of the T&V method, alternative approaches to extension continued to operate in many states, and agricultural research organizations never abandoned their farm demonstration programs.

Similarly, crop-oriented extension programs did not completely disappear, and funding continued to be provided for diffusing veterinary and milk production advice through the cooperative dairy movement. With help from the United States Agency for International Development, the land-grant college approach (which links extension activities to university-based research and training programs) retained influential adherents and scored significant successes, especially in the Uttar Pradesh and Punjab regions of India.

This said, T&V has dominated agricultural extension in South Asia and Africa for more than two decades, partly because of the strong support offered by the World Bank. About 5 percent of the Bank's agricultural lending has been devoted to extension. Currently, sixty-four active Bank-financed projects include extension components, a majority of which apply T&V principles.

Agricultural extension today is at a crossroads. The T&V star has risen high on the firmament of agricultural policy only to become prey to severe criticism and to a new, pluralistic doctrine of agricultural extension (Zijp 1996, World Bank 1997). What explains the rise of an alternative paradigm?

Evaluating T&V Projects

A World Bank study (Purcell and Anderson 1997), based on independent evaluations of thirty-three free-standing agricultural extension projects, shows that [illegible]0 percent had satisfactory outcomes, that is, they met their major relevant objectives efficiently. The share of satisfactory outcomes varied widely across regions—from a high of 83 percent in South Asia to a low of 53 percent in Africa. Although this success rate is higher than that of the Bank's overall agricultural lending portfolio, it compares unfavorably with the performance of the human resource portfolio (about 80 percent).

The T&V approach was used in 90 percent of the projects reviewed. All successful projects helped to heighten the government attention to technology transfer and to increase the volume of work achieved by the implementing agency, resulting in better-trained staff, enhanced coverage of farmers, expanded focus on technology, and improved delivery of extension services. In particular, T&V increased extension agents' contacts with farmers, thanks to staff mobility and the programming discipline associated with the approach.

The study, however, highlighted several disturbing deficiencies.

- Ninety percent of the projects faced budgetary constraints, in part becaus almost half did not evince strong borrower or implementing agenc ownership
- More than half of the projects suffered from inadequate extension messag resulting from research weaknesses or poor linkages between extension an research
- Twenty-five percent of the projects were hindered by the low educatio level of frontline staff
- The training programs of more than half of the projects did not give th frontline staff sufficient practical knowledge, and
- Almost 40 percent of the projects suffered from inadequate adaptation t local conditions.

T&V's hierarchically organized and strictly programmed method of agricul tural extension presumes the availability of a sustained flow of research innova tions coupled with the ability of implementing agencies to secure, retain, an motivate good technical staff. Where both of these elements were availabl T&V may well have accelerated the spread of new agricultural technologies on rewarding scale. Where the initial conditions were not suitable—for instance because farming conditions were highly differentiated, the research pipelin was empty, and either a disciplined organization or adequate skills, or both were lacking—T&V proved poorly adapted to the challenge.

T&V usually has been introduced on a national scale after only limited pilo programs—a pattern that has usually led to sharp increases in budgetary out lays. Antholt (1991) notes that the long-term consequence of increased payroll has had detrimental effects on resource allocations, as nonsalary requirement eventually have been squeezed by the increased emoluments of an aging exten sion cadre. As a result frequent concerns have been expressed regarding th fiscal sustainability of the T&V system.

To be sure, the blueprint nature of extension programming associated wit T&V has been modified to encourage adaptation to local conditions, but th hierarchical mode of operation still lacks flexibility and fails to encourage cos recovery, development of farmer-led programs, or private-sector participation T&V has thus remained dependent on outside sponsorship and support.

In countries with a supply of relevant research innovations and able staff, as in India, "trait making" (that is, the adoption of an imported institutional mode see Hirschman 1967, p. 131) succeeded, and the innovation was integrated into the agricultural administration. Elsewhere (for instance, in Turkey, wher it was first tried), the management discipline of the system eroded, and T&V was eventually abandoned. Remarkably, the Bank study found that only 35

percent of the extension projects with satisfactory outcomes were considered sustainable (Purcell and Anderson 1997).

Thus several Asian countries have found the T&V approach poorly adapted to their needs (Antholt 1994). After a five-year trial of T&V in the early 1980s, Thailand moved to a participatory approach, driven by farmers deciding at the local level what extension service is desired. Malaysia, focusing on tree crops, has emphasized since 1984 a market-driven model, under which farmers contribute to the costs of the extension services received. Neither Bangladesh nor Pakistan was able to induce better extension practices through T&V, and in Indonesia T&V had little impact in dryland, multicrop systems. Currently, T&V's momentum is restricted to Sub-Saharan Africa.

Economic Analysis of Extension Projects

In principle, the economic analysis of extension projects requires systematic comparison of costs and benefits with and without the project (Birkhaeuser, Evenson, and Feder 1991). In practice, systematic social experiments comparing different methods of extension in similarly situated areas have not been conducted. Where extension programs have been evaluated by comparing outcomes in similar contiguous areas, the results have been nuanced. Work by Feder, Slade, and Lau (1985); Feder and Slade (1986); and Feder, Lau, and Slade (1987), which compared productivity differentials in Haryana and Uttar Pradesh in India, suggest that T&V had no significant impact on rice production but yielded economic returns of at least 15 percent in wheat-growing areas. Similar work in Pakistan (Hussain, Byerlee, and Heisey 1994) found smaller effects in wheat areas, although they recorded an increase in the number of contacts between farmers and extension agents, suggesting that the contacts were ineffective.

By contrast, the extraordinarily high rates of return for expenditures on agricultural extension estimated by Bindlish and Evenson in this issue are not robust because of inadequate baseline estimates, uncertainties about the causes of technology adoption, unknown lags in causal effects, and so on.[1]

Of greater relevance to policymakers is the analysis in this volume by Umali-Deininger (1997), which throws light on the appropriate roles of the private, voluntary, and public sectors in funding and delivering agricultural extension services. The paper is a useful exploration of public-funding rationale, as canvassed by Lindner (1993) and recently recommended by Devarajan, Squire, and Suthiwart-Narueput (1997). The implication is clear: where the knowledge being diffused is embedded in or closely associated with market goods (for example, plantation crops, tractors, or hybrid seed), it is best to leave the deliv-

ery of advisory services to the private sector within an appropriate regulatc framework.

Where, however, the technology or practice being promoted is associat with a toll good (such as farm management or marketing information), delive of extension advice is best handled by a judicious combination of public ai private entities (Umali and Schwartz 1994). If a common-pool good is involv (forestry, fisheries, common pastures), it is critical to connect the extensi effort closely to cooperative or voluntary action. Only where market and p ticipation failures are high—for example, where subsistence farming don nates, as it does in Sub-Saharan Africa, or where social conditions preclu voluntary action, as they do in Myanmar—is a pure public-sector approach agricultural extension desirable.

Toward a New Paradigm

Umali-Deininger's thesis reflects the growing influence of the new institutior economics in development thinking (for example, Hoff, Braverman, and Stigl 1993; Picciotto 1995). It offers a pragmatic approach to institutional desig equidistant from the dogmas of massive market failure and cynical "pub choice" theory. The paper may prove seminal, for it provides the theoreti rationale for a pluralistic approach to extension far better adapted to the cu rent challenges of agricultural development than the assembly-line model technology diffusion embodied by T&V. Three challenges underlie the need f a new approach.

First, developing country governments are under severe strain. Not only c these governments ill afford to employ large numbers of extension workers or permanent basis, but their administrative capacities are severely strained by tl demands of an increasingly far-flung and technically sophisticated organizati of extension agents, who they are not always equipped to train, reward, ai motivate (Antholt 1994, p. 28). A new role for the state is emerging that giv pride of place to the creation of enabling environments for private and volu tary action rather than to the direct provision of services.

Second, the perception of agriculture's potential and constraints has change In many situations the dissemination of standard packages of inputs and pra tices is no longer relevant, if indeed it ever was (Simmonds 1988). What increasingly required is an approach that can generate custom-made, enviro mentally friendly solutions based on the farmers' involvement (Anderson 199 Axinn 1991; Eponou 1996; Purcell and Anderson 1997).

Third, the spread of education and modern communications and the rise commercial farming have created opportunities for alliances among the publi

private, and voluntary sectors. More open and liberalized agricultural markets are bringing the knowledge and skills of private agribusiness to farmers without involving public-sector intermediaries. In both more- and less-developed countries, farmer-led approaches to extension are spreading, while farmers' associations, cooperatives, and self-help agencies are contributing handsomely to the diffusion of modern technology.

According to Tendler (1997), informal performance contracts between Brazilian farmers and extension agents have increased the commitment of extension workers, improved the customization of advice, and increased productivity. In Indonesia integrated pest management programs held at the Food and Agriculture Organization's farmer field schools show the value of turning farmers into extension agents and extension agents into farmers, as well as the diffusion potential implicit in group learning and the use of farmers as trainers (Kingsley and Musante 1996).

A total shift from public funding to client funding may not be in the public interest, given the external benefits of technology diffusion and legitimate equity concerns (Dinar 1996). But there are obvious benefits, above all value for money, associated with a demand-driven approach. In some settings public extension systems still need to be involved in the diffusion of technology; in others governments should divest themselves of these support services. Everywhere governments should seek to enhance the voice of farmers and the cost-effectiveness of service delivery. In other words, unbundling the twin government roles of financing extension and actually delivering extension services has become essential.

Centralized mainline extension services must continue to give way to a variety of hybrid solutions combining public support with private delivery methods. Cost sharing and voucher systems can increase the voice of farmers in the management of extension systems (World Bank 1990; Antholt 1994). Contract extension, where extension agents contract with farmers to provide the information they request (long practiced in China), can increase responsiveness. Ecuador extension agents sharecrop with farmers for a profit. Costa Rica has experimented with vouchers that promote private technical assistance to small- and medium-scale producers.

Similarly Chile publicly finances 70 percent of the costs of private technology-transfer firms, which contract with small-scale producers; similar services operate in Mexico and Venezuela. In New Zealand gradually rising cost-recovery targets were set for the public extension service and were easily exceeded from 1988 until 1994, at which time the service was profitable and the operation was privatized (Milligan 1997). Finally the successful introduction of cost recovery for extension services on a pilot scale in Nicaragua, described in this volume by Keynan, Olin, and Dinar (1997) confirms the feasibility of a demand-driven

approach directed to smallholders and the bracing impact of pricing on service standards.

Conclusion

In an effort to contribute effectively to the well-being of rural areas in developing countries, agricultural extension organizations are adjusting to a new order characterized by less government funding and more differentiated requirements. Agricultural extension will increasingly rely on new information technologies and multiple knowledge networks involving the private sector and civil society. There are many ways to gradually divest the public sector from its extension activities. Workable and sustainable extension systems call for tailor-made experimentation informed by global experience.

Accordingly, rather than pressing governments for increased budgetary allocations for public-sector extension systems, development assistance agencies should support policies aimed at increasing the role of users, private companies and the voluntary sector and should assist governments in enhancing the cost effectiveness and quality of existing services through institutional innovation and outsourcing. This approach would allow the public sector to concentrate its limited resources on providing services to neglected areas and high-leverage actions directed at education and training, information technology, and the creation of enabling frameworks for equitable and environmentally sustainable rural development.

Notes

Robert Picciotto is director general of operations evaluation in the World Bank Group. Jock R. Anderson is evaluation adviser in the Operations Evaluation Department. The authors are grateful to Charles Antholt, Daniel Benor, Vishva Bindlish, Hans Binswanger, John English, Robert Evenson, Gershon Feder, Madhur Gautam, Jacob Meerman, Michel Petit, Roger Slade and Willem Zijp for their comments.

1. In experimenting with alternative specifications using the authors' data from Kenya, minor adjustments were found to cause radically different implied extension effects, indicating a lack of robustness in the key findings.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Anderson, Jock R. 1991. "FSRE [Farming Systems Research and Extension] Impact Inquisition: Investor Issues." *Journal of the Asian Farming Systems Association* 1(1):55–68.

Anderson, Jock R., and Karla Hoff. 1993. "Technological Change, Imperfect Markets, and Agricultural Extension: An Overview." In Karla Hoff, Avishay Braverman, and Joseph E. Stiglitz, eds., *The Economics of Rural Organization: Theory, Practice, and Policy.* New York: Oxford University Press for the World Bank.

Antholt, Charles H. 1991. "Agricultural Extension in the 21st Century: Lessons from South Asia." In W. M. Rivera and D. J. Gustafson, eds., *Agricultural Extension: Worldwide Institutional Evolution and Forces for Change.* Amsterdam: Elsevier.

—. 1994. *Getting Ready for the Twenty-First Century: Technical Change and Institutional Modernization in Agriculture.* World Bank Technical Paper 217. Washington, D.C.

Axinn, George H. 1991. "Potential Contribution of FSRE to Institution Development." *Journal of the Asian Farming Systems Association* 1(1):69–78.

Axinn, George H., and Sudhakar Thorat. 1972. *Modernizing World Agriculture: A Comparative Study of Agricultural Extension Education Systems.* New Delhi: Oxford & IBH Publishing Co.

Baxter, Michael, Roger Slade, and John Howell. 1989. *Aid and Agricultural Extension. Evidence from the World Bank and Other Donors.* World Bank Technical Paper 87. Washington, D.C.

Benor, Daniel, James Q. Harrison, and Michael Baxter. 1984. *Agricultural Extension: The Training and Visit System.* World Bank, Washington, D.C.

Bindlish, Vishva, and Robert E. Evenson. 1997. "The Impact of T&V Extension in Africa: The Experience of Kenya and Burkina Faso." *The World Bank Research Observer* 12(2):183–201.

Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder. 1991. "The Economic Impact of Agricultural Extension: A Review." *Economic Development and Cultural Change* 39(3):607–50.

[illegible] Shantayanan [illegible] Squire, and [illegible] 1997. "Beyond Rate of Return: Reorienting Project Appraisal." *The World Bank Research Observer* 12(1):35–[illegible]

[illegible] 1996. "Extension Commercialization: How Much to Charge for Extension Services." *American Journal of Agricultural Economics* 78(1):1–12.

[illegible]son, Graham [illegible] 1997. "Government-Sponsored Rural Development: Experience of the World Bank." In C. Peter Timmer, ed., *Agriculture and the State: Growth, Employment, and Poverty in the Developing Countries.* Ithaca, N.Y.: Cornell University Press.

[illegible] 1996. *Partners in Technology Generation and Transfer: Linkages between Research and [illegible] Organizations in Three Selected African Countries.* Research Report 9. International Service for National Agricultural Research, The Hague.

[illegible] Robert E., Carl E. Pray, and Mark W. Rosegrant. Forthcoming. *Agricultural Research and Productivity Growth in India.* Washington, D.C.: International Food Policy Research Institute.

[illegible] Gershon, Lawrence J. Lau, and Roger H. Slade. 198[illegible] "Does Agricultural Extension [illegible] The Training and Visit System in Northwest India." *American Journal of Agricultural Economics* 69(3):677–86.

[illegible] Gershon, and Roger Slade. 1986. "The Impact of Agricultural Extension: The Training and Visit System in India." *The World Bank Research Observer* 1(2):139–61.

[illegible] Gershon, Roger Slade, and Lawrence Lau. 1985. *The Impact of Agricultural Extension: The Training and Visit System in Haryana.* World Bank Staff Working Paper 756. Washington, D.C.

Feder, Gershon, and Dina L. Umali. 1993. "The Adoption of Agricultural Innovations." *Technological Forecasting and Social Change* 43:215–39.

Foster, Andrew D., and Mark K. Rosenzweig. 1996. "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution." *American Economic Review* 86(4):931–53.

Hirschman, Albert O. 1967. *Development Projects Observed.* Washington, D.C.: Brookings Institution.

Hoff, Karla, Avishay Braverman, and Joseph E. Stiglitz, eds. 1993. *The Economics of Rural Organization: Theory, Practice, and Policy.* New York: Oxford University Press for the World Bank.

Hussain, Syed S., Derek Byerlee, and Paul W. Heisey. 1994. "Impacts of the Training and Visit Extension System on Farmers' Knowledge and Adoption of Technology: Evidence from Pakistan." *Agricultural Economics* 10(1):39–47.

Keynan, Gabriel, Manuel Olin, and Ariel Dinar. 1997. "Cofinanced Public Extension in Nicaragua." *The World Bank Research Observer* 12(2):225–47.

Kingsley, M. A., and P. Musante. 1996. "Activities for Developing Linkages and Cooperative Exchange among Farmers' Organisations, NGOs, GOs, and Researchers: Case Study of an NGO-Coordinated Integrated Pest Management Project in Indonesia." In Vanessa Scarsborough, ed., *Farmer-Led Approaches to Extension. Papers Presented to a Workshop in the Philippines, July 1995.* London: Overseas Development Institute.

Klitgaard, Robert. 1995. *Institutional Adjustment and Adjusting to Institutions.* World Bank Discussion Paper 303. Washington, D.C.

Lindner, Robert K. 1993. "Privatising the Production of Knowledge: Promise and Pitfalls for Agricultural Research and Extension." *Australian Journal of Agricultural Economics* 37(3):205–25.

Lipton, Michael, with Richard Longhurst. 1989. *New Seeds and Poor People.* Baltimore, Md.: The Johns Hopkins University Press.

Macklin, Michael. 1992. *Agricultural Extension in India.* World Bank Technical Paper 190. Washington, D.C.

Meerman, Jacob. 1997. *Reforming Agriculture. The World Bank Goes to Market.* World Bank Operations Evaluation Study. Washington, D.C.

Milligan, Keith E. 1997. *Extension and Technology Transfer in New Zealand, 1970–1996.* Lincoln International Ltd., Canterbury, New Zealand.

OED (Operations Evaluation Department). 1988. *Rural Development. World Bank Experience 1965–86.* World Bank Operations Evaluation Study. Washington, D.C.

Pasour, E. C., Jr. 1990. *Agriculture and the State: Market Processes and Bureaucracy.* New York: Holmes & Meier for the Independent Institute.

Picciotto, Robert. 1995. *Putting Institutional Economics to Work: From Participation to Governance.* World Bank Discussion Paper 304. Washington, D.C.

Raman, K.V., T. Balaguru, and P. Manikandan. 1988. *National Agricultural Research, Education, and Extension Education Systems in India.* Hyderabad, India: National Academy of Agricultural Research Management.

Schwartz, Lisa A., and Jacob Kampen. 1992. *Agricultural Extension in East Africa.* World Bank Technical Paper 164. Washington, D.C.

Simmonds, Norman W. 1988. "Observations on Induced Diffusion of Innovations as a Component of Tropical Agricultural Extension Systems." *Agricultural Administration and Extension* 28(2):207–16.

Tendler, Judith. 1997. *Good Government in the Tropics.* Baltimore, Md.: The Johns Hopkins University Press.

Timmer, C. Peter. 1991. "The Role of the State in Agricultural Development." In C. P. Timmer, ed., *Agriculture and the State: Growth, Employment, and Poverty in the Developing Countries.* Ithaca, N.Y.: Cornell University Press.

Umali-Deininger, Dina. 1997. "Public and Private Agricultural Extension: Partners or Rivals?" *The World Bank Research Observer* 12(2):203–24.

Umali, Dina L., and Lisa Schwartz. 1994. *Public and Private Agricultural Extension: Beyond Traditional Frontiers.* World Bank Discussion Paper 236. Washington, D.C.

World Bank. 1990. *Agricultural Extension: The Next Step.* Washington, D.C.

———. 1997. "From Vision to Action in the Rural Sector." Washington, D.C.

Zijp, Willem. 1994. *Improving the Transfer and Use of Agricultural Information: A Guide to Information Technology.* World Bank Discussion Paper 247. Washington, D.C.

———. 1996. "Promoting Pluralism." Presentation to the Extension Workshop: Alternative Mechanisms for Funding and Delivering Extension. Agriculture and Rural Development Department, Washington, D.C. Processed.

Creating a Legal Framework for Economic Development

Richard A. Posner

A modernizing nation's economic prosperity requires at least a modest legal infrastructure centered on the protection of property and contract rights. The essential legal reform required to create that infrastructure may be the adoption of a system of relatively precise legal rules, as distinct from more open-ended standards or a heavy investment in upgrading the nation's judiciary. A virtuous cycle can arise in which initially modest expenditures on law reform increase the rate of economic growth, in turn generating resources that will enable more ambitious legal reforms to be undertaken in the future.

It used to be common in the economic literature on development to enumerate the multitudinous sources of market failure and prescribe complex government interventions to cure them without paying much attention to the equally numerous sources, especially in poor countries, of governmental failure (Stiglitz 1994). A 1995 conference on legal reform, as well as other recent judicial reform initiatives (Rowat, Malik, and Dakolias 1995; Dakolias 1996), attest to the growing awareness that the failure of governments in poor countries to provide the basic framework of a capitalist economy may be an important factor in keeping poor countries poor. Markets are more robust than some market-failure specialists believe. But their vigor may depend on the establishment of an environment in which legal rights, especially property and contractual rights, are enforced and protected—an environment that is taken for granted in wealthy nations (see Gray 1991 and Rausser 1992 for good statements of this essentially Weberian point).

The citizens of wealthy countries take this legal machinery for granted because it works well enough most of the time and because it does not cost a great deal. In its ideal form (an important qualification), the machinery consists of competent, ethical, and well-paid professional judges who administer rules that are well designed for the promotion of commercial activity. The judges are insulated from interference by the legislative and executive branches of government. They are advised by

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 1–11

point out that it is more costly and time-consuming to create efficient legal institutions than to enact efficient rules for the existing inefficient institutions to administer. The creation and dissemination of a rule involves small fixed costs and (like other information goods) negligible marginal costs, while legal institutions require heavy inputs of high-priced, educated labor. This implies that the rules-first strategy is more advantageous in more populous countries because the average costs are lower. China, the most populous poor country, followed this approach by introducing modern, commercially oriented rules of law at the same time that it liberalized the economy (Potter 1994).

RULES FIRST. This strategy can serve as the starting point of several short-term measures for improving legal institutions as well. It is important to emphasize that such measures be both efficient and, less obviously, rules, and to distinguish between *substantive* and *procedural* efficiencies. A rule is substantively efficient if it sets forth a precept that internalizes an externality or otherwise promotes the efficient allocation of resources: a rule forbidding the use of another person's property without consent is an example. A rule is procedurally efficient if it is designed to reduce the cost or increase the accuracy of using the legal system.

Examples are: a requirement that contracts be in writing to be legally enforceable; a rule that no claim of infringement of legal rights is enforceable unless filed within three years of the alleged infringement; a requirement that certain disputes, say between employers and employees or between securities brokers and their customers, must be referred to binding arbitration; and a rule entitling the winner of a judgment for damages to receive interest on the judgment at the market rate from the date the suit was filed. The first two of these rules are designed to reduce the information costs of the legal system, the third to reduce the judicial workload by shunting certain disputes to an alternative method of dispute resolution, and the last to enable judges to use delay to cope with a heavy workload without destroying the utility of the legal system to persons whose rights have been infringed.

The rule requiring arbitration has additional importance as a method of encouraging the formation of trade associations and other business groups, useful intermediaries between the state and the individual or family in a commercial society (Casella 1996). Lawyers tend to ignore such procedural rules, even though they consider themselves experts on procedure, and to emphasize legal doctrine at the expense of methods that actualize it. The procedural rules listed here have received much less attention from lawyers than have the intricacies of doctrine.

ADMINISTRATIVE CONSIDERATIONS. I want to make the point that these are "rules" rather than "standards" because determining whether they have been violated is a relatively mechanical, cut-and-dried process rather than one requiring the exercise of discretion or the determination of numerous facts. The trespass rule requires

determining, for instance, where the boundary line is, whether it has been crossed, and whether there was consent; a statute of limitations requires determining only the date on which the alleged infringement occurred and the date on which the suit was brought; and so on. Lawyers being what they are, the actual administration of rules is more complex than I am letting on. But it is simpler than the administration of standards, such as negligence, bad faith, unreasonable restraint of trade, and unconscionability—at least if professional judges rather than lay jurors are used to determine violations. This is an important qualification. Standards, to the extent that they are intuitive (such as the concept of due care, which is the heart of the negligence standard), may be easier to understand, accept, and apply than rules, which may be simplified to the point of arbitrariness, as in the case of statutes of limitations. Because understanding and acceptance are important to achieving voluntary compliance as well as to the sensible decision of cases by lay adjudicators, a mixture of rules and standards is optimal. But in poor countries with weak legal traditions, the tilt should be in favor of rules because they are easier to administer.

The relative simplicity of rules has two consequences for the kind of weak judiciary that one is apt to find in a poor country. The first is that the application of rules places fewer demands on the time and the competence of the judges and is therefore both cheaper and more likely to be accurate. The accuracy is a little illusory, because it is a property of governance by rules that they never quite fit the complex reality that they govern. But this observation is consistent with their being more efficient than standards if administered by a judiciary that has a limited capability for the kind of nuanced and flexible decisionmaking that standards require. Second, rules facilitate monitoring of the judges and so reduce the likelihood of bribery and the influence of politics in the judicial process. The less discretion a judge has in making decisions, the easier it will be to determine whether a case has been decided contrary to law or whether there is a pattern of favoring one class or group of litigants over another.

Adoption of Foreign Laws

The adoption of such a set of rules is much more easily said than done, but perhaps not so much more easily. There is, to begin with, a long tradition of what is called the "reception" of foreign law. When the American colonies broke away from Great Britain, each new state decided how much of the common law of England would be received into and made a part of the law of the state (Wood 1969). This pattern is typical in former colonies; for example, half a century after independence, the Indian legal system retains the strong imprint of its English origins. But the reception of foreign law is not limited to cases of original imposition. The Japanese and Chinese also borrowed extensively from European legal codes (Ma 1995; Ford 1996). After

World War II the occupying powers imposed a variety of legal changes on Germany and Japan, for example in competition law and (in Japan) criminal procedure, that have taken root. European Community law is becoming incorporated into the domestic law of the member nations.

Such grafts do not always take: a notable example was the adoption by most South American countries in the nineteenth century of constitutions modeled on that of the United States. Some of the Western-inspired constitutions in the former communist nations of Central and Eastern Europe may encounter the same fate. But constitutional law is a special case; its effectiveness depends on a particularly complex cultural and institutional matrix. In other areas, the prospects for transplanting Western law are good.

A poor country might do worse than adopt portions of the U.S. Uniform Commercial Code (or its European counterpart), a simple and successful set of rules and standards governing primarily sales of goods, negotiable instruments such as checks and letters of credits, and secured transactions such as mortgages. The poor country might want to modify or delete some of the standards in the Code, such as the standard of good faith, and to require (assuming that illiteracy is not widespread) that more contracts be in writing. And it might want to extend the Code, again with various modifications, to transactions that in the United States are left to the common law, such as contracts involving the sale of services rather than goods—for example, insurance and construction contracts.

This proposal has, I admit, a somewhat roundabout character. Historically, commercial law originated in the customs of merchants enforced privately by arbitration or equivalent informal methods and only later adopted by the courts. The Uniform Commercial Code is in part an attempt to codify commercial practices. To the extent that the business community in a poor country has its own law, it may be better to codify that law than to try to borrow another country's model. But the law may be underdeveloped in a poor country—may in fact be a cause of underdevelopment or a symptom, or both—and the task of codification may require technical skills of drafting and organization that are in short supply. In these circumstances the adoption of a foreign code may be the more sensible move. The important point is that both foreign law and the application of local custom as formal law are well-tried methods by which a nation can adopt a legal code without starting from scratch—or needing a Napoleon.

It is important, however, to adapt the imported code to the local culture (Rubin 1994), a task for local, not foreign, lawyers, who know something about the country whose law they are borrowing. I do not advise dispatching European or American lawyers to tell a country how to adapt foreign laws to its legal and social institutions and stage of economic development. One approach might be to establish a law reform commission to rationalize, unify, and modernize national laws, borrowing wherever possible from established foreign models.

The Judiciary

The fundamental tradeoff is between making a rather modest investment in better rules and making a big investment in the judiciary. And it is a tradeoff. If judicial salaries are high enough and tenure sufficiently secure, the judiciary of even a poor country will be able to attract competent and honest lawyers. But highly educated people, who are needed to staff a good court and also to appear before it as advocates, are a very scarce resource in poor countries, making the opportunity costs of a first-rate judiciary and its associated bar of practitioners very high. And if the salaries of one class of officials far exceed those of other officials, it could create a ripple effect throughout the entire civil service, resulting in large fiscal costs and a large drain on the limited talent of the nation. Finally, the political authorities will be reluctant to create a corps of truly independent officials who may constitute a rival center of power; or they may lack the power to protect the judges from private violence if they stand up to powerful interest groups. Moreover, if the political authorities are weak or corrupt, generous compensation may simply increase the value of a judgeship as a patronage plum and result in an actual decrease in judicial quality.

The more costly it is to create a high-quality independent judiciary, the more beneficial it is to focus legal reform on the adoption of substantively and procedurally efficient rules. In emphasizing this point, however, I do not propose to abandon entirely the task of improving legal institutions. Indeed, if the law's administrative infrastructure is sufficiently weak, even good rules may simply be ignored. This appears to be the case in Russia, which has several modern legal codes on its books. Although a regime of rules reduces the likelihood of financial or political corruption, rules cannot be the complete answer. Countries may also need to alter the structure of judicial salaries. Specifically, the more that judicial compensation is "backloaded" in the form of generous pension rights that are forfeited if the judge is removed from office for incompetence or venality, the greater the incentive of the judge to behave with integrity (Becker and Stigler 1974). If the cost is very great, even if the likelihood of being detected is slight, the appointee may be deterred. Another corruption-fighting incremental change worth considering is having judges sit in panels—or with juries—rather than by themselves, to increase the transaction costs of bribery and the likelihood of discovery. Unlike the compensation adjustment, this proposal would cost something because more judges would be needed—even if lay juries were used—since jury trials take longer.

Issues of Enforcement

Where the suggested approach emphasizing rules over institutional reform is most likely to fall short is in securing people against the threat of government confiscation. It is all very well to have well-defined private property rights determine legal rem-

edies against the invasion of those rights by private parties and to devise an effective system of contract rights to exchange property among private owners and entrepreneurs. But these rights may mean little if the state can seize the fruits of successful investment (North and Weingast 1989). Although certain rules, if enforced, will prevent this—rules such as forbidding the state from taking property without just compensation or forbidding discriminatory taxation—their efficacy depends on the willingness of the judges to stand up to government officials. It looks as though we are back to needing the competent, ethical, well paid, and politically independent judiciary so unlikely to be feasible for a poor nation.

A solution to this dilemma may be to establish a special court, as in France, whose sole mission is to restrain the government, the *Conseil d'État* (Merryman 1996). The judges of this court must be competent, ethical, and well paid, but because the court's jurisdiction will be so circumscribed, the resources required to equip it will be modest. If the court is confined to purely economic issues, moreover, the political authorities may be willing to tolerate its independence, especially if they understand its importance for the economy. An alternative may be to turn over judicial authority to a regional or international court, although enforcing the decrees of that court against the government of the country may be problematic.

Issues of Commitment

It should be obvious that effective legal reform depends ultimately on a political will to reform, which in turn is likely to depend on a political will to implement economic reform. If the dominant political groups in society want economic prosperity and are willing to risk the loss of political control over the economy that modern economic conditions dictate, they will also want legal reform. If they do not want economic reform, the will to adopt legal reform is likely to be absent.

I emphasize the importance of modest fiscal outlays in creating a virtuous cycle of legal and economic reform. Remember that economic progress is possible with little perhaps with no—law and can be stifled by excessive investment in public-sector projects, including legal reform. A small expenditure on law reform can increase the rate of economic growth, which will in turn generate additional resources for more ambitious legal reforms later.

Criminal Law

I have talked only about property and contract rights, and fleetingly about concepts such as trespass that back up property rights. I have said nothing about the criminal laws or human rights, which are often reciprocals: many basic human rights protect citizens against the overvigorous enforcement of criminal laws. It is sometimes ar-

gued that economic rights are inseparable from political rights. Drèze and Sen 1995 have pointed out that countries with a free press do not suffer from famines; and the right to vote and the right of free speech place constraints on government that may in turn protect business interests against the risk of confiscatory measures. I am not very confident about the empirical significance of these points, however. Given the role of technology, it is hard to imagine a country concealing a major famine even if its own media are gagged; North Korea has certainly failed. And democracy may not do much for the economy; India has not been economically more progressive than the nondemocratic nations of Asia.

As for granting extensive rights to criminals, this is bound to undermine the efficacy of the criminal laws, and by doing so, unsettle property rights. Rights make it harder to convict the guilty as well as the innocent. Sophisticated police forces and prosecutors can apprehend and convict the guilty without trampling on rights; but sophisticated law enforcement is costly. This point is especially salient in countries such as Russia where acquisitive crime is so rampant that it retards economic development. In such countries a strict criminal law and a corresponding de-emphasis on the protection of civil liberties may be an important part of legal reform and an important tool for the protection of property and contract rights.

So my message is a modest, and perhaps a harsh, one. Legal reform is an important part of the modernization process of poor countries, but the focus of such reform should be on creating substantive and procedurally efficient rules of contract and property rather than on creating a first-class judiciary or an extensive system of civil liberties. This is a general prescription, however, and the proper legal structure for an individual country will depend on a host of considerations that I have not attempted to canvas in this overview.

Notes

Richard A. Posner is chief judge, United States Court of Appeals for the Seventh Circuit and Senior Lecturer, University of Chicago Law School. This article is based on an address given at the World Bank Workshop on Legal Reform on April 14, 1997. I thank Cheryl Gray for inviting me to give the address, Sorin Feiner and Andrew Trask for very helpful research assistance, and Stephen Holmes and three anonymous reviewers for the *World Bank Research Observer* for many helpful comments on a previous draft.

1. This may explain the negative correlation across countries between the number of lawyers and the rate of economic growth (Murphy, Shleifer, and Vishny 1991). The correlation is misleading because much of the output of lawyers consists of nonmarket goods; but these may not be as important in poor countries as in wealthy ones.

2. An employee with firm specific skills is likely to have a better job at this firm than at a different one. Such an employee will be reluctant to quit and the employer will be reluctant to dispense with his or her skills.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 106:407–43.

Becker, Gary S. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* 3d ed. Chicago: University of Chicago Press.

Becker, Gary S., and George J. Stigler. 1974. "Law Enforcement, Malfeasement, and Compensation of Enforcers." *Journal of Legal Studies* 3(1):1–18.

Casella, Alessandra. 1996. "On Market Integration and the Development of Institutions: The Case of International Commercial Arbitration." *European Economic Review* 40(1):155–86.

Dakolias, Maria. 1996. *The Judicial Sector in Latin America and the Caribbean: Elements of Reform.* Technical Paper 319. Washington, D.C.: World Bank.

Dreze, Jean, and Amartya Sen. 1995. *India: Economic Development and Social Opportunity.* New Delhi: Oxford University Press.

Ford, Christopher A. 1996. "The Indigenization of Constitutionalism in the Japanese Experience." *Case Western Reserve Journal of International Law* 28(1):3–62.

Gray, Cheryl W. 1991. "Legal Process and Economic Development: A Case Study of Indonesia." *World Development* 19(7):763–78.

———. 1997. "Reforming Legal Systems in Developing and Transition Economies." *Finance and Development* 94(3).

Hay, Jonathan R., Andrei Shleifer, and Robert W. Vishny. 1996. "Privatization in Transition Economies: Towards a Theory of Legal Reform." *European Economic Review* 40(315):559–68.

Hendley, Kathryn, Barry W. Ickes, and Randi Ryterman. 1997. "Observations on the Use of Law by Russian Enterprises." *Post-Soviet Affairs* 13(1):19–41.

Intriligator, Michael D. 1994. "Privatisation in Russia Has Led to Criminalisation." *Australian Economic Review* 105(April–June):4–14.

Ma, Herbert H. P. 1995. "The Chinese Concept of the Individual and the Reception of Foreign Law." *Journal of Chinese Law* 9(2):207–21.

McGrory, Daniel. 1995. "Civilizing the Russian Underground Economy: Requirements and Prospects for Establishing a Civil Economy in Russia." *Transnational Law and Contemporary Problems* 5(1):65–98.

Merryman, John Henry. 1996. "The French Deviation." *American Journal of Comparative Law* 44:109–16.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1991. "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics* 106(2):503–30.

North, Douglass C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History* 49(4):803–32.

Posner, Richard A. 1996. "The Cost of Rights: Implications for Central and Eastern Europe and for the United States." *Tulsa Law Journal* 32(1):1–19.

———. Forthcoming. "Equality, Wealth, and Political Stability." *Journal of Law, Economics, and Organization.*

Potter, Pitman P. 1994. "Riding the Tiger: Legitimacy and Legal Culture in Post-Mao China." *China Quarterly* 138:325–58.

Rausser, Gordon C. 1992. "Lessons for Emerging Market Economies in Eastern Europe." In Christopher C. Clague and Gordon C. Rausser, eds. *The Emergence of Market Economies in Eastern Europe.* Cambridge, MA: Blackwell Publishers.

Rowat, Malcolm, Waleed H. Malik, and Maria Dakolias, eds. 1995. *Judicial Reform in Latin America and the Caribbean: Proceedings of a World Bank Conference.* Technical Paper 280. Washington D.C.: World Bank.

Rubin, Paul H. 1994. "Growing a Legal System in the Post-Communist Economies.' In *Cornell International Law Journal* (27)1:1–47.

Scully, Gerald W. 1988. "The Institutional Framework and Economic Development." *Journal of Political Economy* 96(4):652 62

Stiglitz, Joseph E. 1994. "The Role of the State in Financial Markets." In *Proceedings of the World Bank Annual Conference on Development Economics 1993.* Washington, D.C.: World Bank.

Wood, Gordon S 1969. *The Creation of the American Republic 1776–1787* New York: Norton.

The Domestic Benefits of Tropical Forests: A Critical Review

Kenneth M. Chomitz • Kanta Kumari

Many forest conservation projects seek to preserve biodiversity by protecting habitats from exploitation or degradation. Although such efforts are often motivated by global concerns, habitat protection also yields domestic benefits. Some of these are intangible or difficult to quantify; others, such as watershed protection and the production of nonforest timber products, are immediate and tangible.

There are two rationales for quantifying the domestic benefits of habitat conservation. The first is motivational. Host countries capture only a small proportion of the global benefits which stem from biodiversity conservation. Demonstration of palpable local benefits could help to build support for biodiversity-oriented projects. Second, the magnitude of domestic benefits could influence project financing. Sufficiently large net domestic benefits could justify financing of a project on narrow economic grounds, with biodiversity conservation as a by-product.

This review finds that the quantifiable benefits of forest preservation in providing hydrological services and nontimber forest products are highly variable. Locally important in some situations, these classes of domestic benefits may in general be smaller than popularly supposed. This underscores the need for financing conservation from the Global Environmental Facility or other global sources rather than placing the burden entirely on domestic resources.

This article focuses almost exclusively on forests in the humid tropics and on two of their potentially most important benefits: hydrological benefits such as erosion control and regulation of stream flows; and nontimber forest products, such as rubber, rattan, fruit, and nuts. Hydrological effects are emphasized for three reasons. First, forests are assumed to be economically important for preventing soil erosion and flooding, protecting the water supply, and maintaining rainfall patterns (Botkin and Talbot 1992, p. [illegible]1; Myers 1995). Second, these assertions are often made with little supporting

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 13–35

evidence. The claims are often seen as a priori plausible or even obvious, although the scientific literature has been questioning some elements of this received wisdom for at least a decade (Hamilton and King 1983). Only a handful of economists have attempted to measure the value of these hydrological functions, and this small literature appears not to have been fully integrated with the scientific literature. Finally, hydrological impacts are potentially of great interest for domestic policy because they involve local externalities: upslope actions affect downslope populations. Also of interest are nontimber forest products, which are increasingly seen as a source of domestic benefits. The emergence of a more extensive literature on nontimber forest products offers an opportunity to assess the valuation of these benefits.

Two themes guide this review. Although these themes are not novel and in fact appear to be quite obvious, we believe that they have been insufficiently emphasized in the existing literature.

First, *benefits must be computed relative to an alternative land use.* The literature tends to treat the benefits of habitat protection as an absolute number irrespective of alternative uses of the land. This is ecologically and economically invalid. Hydrological functions of the land are strongly related to ground cover, as we discuss at length. Hence, the hydrological impact of converting a natural forest to a plantation might be quite different from converting it to annual cropping, and this will affect the value of maintaining the land as forest.

More generally, we are interested in protecting areas precisely because they are in current or future danger of being converted to an alternative use. Therefore, to argue that a particular area should remain protected for economic reasons, the benefit stream provided by the forest must be compared with the benefit stream that would result from the likely alternative. In other words, it is necessary to compute the net benefits of forest preservation: the gross benefits under protection less the forgone benefits from the alternative use (opportunity cost).

Second, *benefit levels are highly location specific and scale dependent.* Habitats in general—and forests in particular—are internally quite heterogeneous. Any sizable forest area is likely to contain many varieties and densities of species, types of soil and terrain, and areas that are more or less accessible to markets. This diversity in turn results in a continuous variation over the landscape in both the physical processes underlying forest benefits and in their economic value. For example:

- The value of forest products depends on the density of the valuable species and on the cost of transportation from the extraction site to the consumer.
- A forest's recreational value depends on its views, accessibility, and species mix.
- Its hydrological value depends on the slope, rainfall, type of soil, position in the watershed, and proximity to dams, fisheries, and irrigation systems.
- The opportunity costs of forest preservation depend on how accessible the land is to markets, and the suitability of the soil for crops.

Moreover, several hydrological processes are scale dependent: the dynamics of erosion and runoff, for instance, are quite different in 100-, 10,000-, and 1,000,000-hectare watersheds. Scale also affects markets for nontimber forest products: product prices may decline as the supply increases. As a result the values estimated for a small site cannot easily be extrapolated to a large area; simple scaling-up of site-specific estimates will yield inaccurate, and often biased, results.

The Hydrological Benefits

Conversion of forest land to other uses can disturb the functioning of the forest ecosystem and ultimately its economic value. (This section draws heavily on Bruijnzeel 1990, and Hamilton and King 1983.) Table 1 shows how changes in land use that affect hydrology are linked to economic impacts.

The first set of links involves the effect of changes in land use on river sedimentation. Three questions are evaluated: First, under what conditions does deforestation increase erosion? Second, what is the relation between increased erosion and delivery of the resulting sediment to downstream economic activities? And third, what is the relation between sediment delivery and subsequent economic damage to dams, canals, harbors, and fisheries?

From Land-Use Change to Sedimentation

There are two links here: from land-use change to erosion, and from erosion to sedimentation. The first is relatively well understood from experimentation and observation on relatively small plots, although most attention focuses on surface erosion as opposed to erosion that causes gullies and mass wasting (landslides). A review

Table 1 *Hydrological-Economic Links*

Hydrological changes	*Economic impacts*
Increased sediment delivery	Siltation of reservoirs, canals, harbors
	Damage to fisheries
	Improved agricultural productivity from downslope soil depletion
Erosion	Loss of productivity for downslope farmers
Increased water yield (runoff and subground flows)	Flood damage to crops and settlements
	Benefits to downstream water consumers
Changes in the water table	Agricultural productivity and household water consumption
Climate change	Impacts on agricultural productivity

Source: Authors

of the best available summary of eighty studies on erosion is quite striking (Wiersum 1984, reproduced in Bruijnzeel 1990, p. 117). Ground cover, rather than canopy, is the chief determinant of erosion. Erosion rates are low in natural forests and in tree gardens, in the fallow phase of slash-and-burn cultivation, and in plantations where weeds and leaf litter are retained. Erosion rates in plots under current slash-and-burn cultivation are ten times as high as in natural forest. In plantations where weeds and litter have been removed, erosion is more than a hundred times as great as in natural forests (table 2).

In many cases erosion may result from road construction associated with logging in the forest rather than from a change in land use. For instance, Hodgson and Dixon (1988) find that the rate of erosion in Palawan, the Philippines, increased four times as a result of logging, but the conversion of uncut forest to road surface increased erosion by a factor of 260. Thus, although roads accounted for only 3 percent of the surface area in the area studied, they were estimated to account for 84 percent of the surface erosion.

Gully erosion and mass wasting are also important sources of sediment, but these processes are more complex than sheet erosion, and less is known about them. Still, it seems fair to conclude that little sedimentation-related damage results from converting natural forests to appropriately managed plantations, agroforestry, moderate grazing, and shifting cultivation with long rotation periods. Road construction, annual cropping, and plantations that remove litter can generate considerable erosion, however. To assess the potential damage, we turn to the next question: Will surface erosion induced by a change in land cover result in major increases in sedimentation?

Table 2. *Relation between Land Cover and Erosion*
(tons per hectare per year)

Type of land cover	*Surface erosion* *Minimum*	*Median*	*Maximum*
Natural forests	0.03	0.3	6.2
Shifting cultivation, fallow period	0.05	0.2	[illegible]
Forest plantations, undisturbed[a]	0.02	0.6	6.2
Multistoried tree gardens[b]	0.01	0.1	0.15
Tree crops with cover crop/mulch	0.1	0.8	5.6
Shifting cultivation, cropping	0.4	2.8	70
Agricultural intercropping in young forest plantations	0.6	5.2	17.4
Tree crops, clean-weeded	1.2	48	183
Forest plantations, litter removed or burned	5.9	53	105

a. Refers to forests for timber production, as opposed to tree crops.

b. A system in which various perennial and sometimes a few annual crops are cultivated simultaneously with trees.

Source: Wiersum (1984), reproduced in Bruijnzeel (1990), p. 117.

The answer depends on two factors. First, only a portion of the eroded soil makes its way into rivers and streams; the remainder is trapped (perhaps temporarily) downslope. The amount of sediment deposited varies inversely with the size of the catchment basin; larger basins have more places for the sediment to get caught than do smaller ones. Mahmood (1987) suggests that the sediment delivery ratio—the proportion of eroded material in a watershed that is carried by a stream—declines from almost 100 percent in basins measuring 200 square hectares to about 10 percent in basins of a million square kilometers. Sediment delivery ratios tend to be about 0.3 in basins measuring hundreds of square kilometers. And lower ratios are associated with longer sediment transport times, causing a lag between the change in the land cover and the downstream impacts.

Second, the induced sedimentation may be large or small relative to existing, or background, sedimentation levels, which vary depending on local geology and the current state of land use in the basin.[1] Background sedimentation is related to existing agriculture and the configuration of roads within a catchment basin and to unstable river banks, natural landslides, and commercial dredging for sand and gravel (Enters 1992; Bruijnzeel 1989, 1990). In general, background sedimentation levels are underestimated because sampling rates are too low to capture infrequent but highly erosive episodes (Mahmood 1987; Bruijnzeel 1990). When this bias is not recognized, higher-than-expected siltation rates at new dams are sometimes erroneously attributed to contemporaneous changes in land use.

Given the complexity of erosion and sediment transport processes and their sensitivity to biological and geological conditions, is it possible to calibrate the relation between changes in land use and the amount of sediment deposited in a watershed? One approach models erosion and transport over the watershed, using mapped data on precipitation, land cover, and topography. This approach has been used to derive the universal soil loss equation for temperate locations, a simple formula based on land cover, precipitation, and slope. It is generally poorly calibrated, especially for tropical areas, and its use is often criticized. Researchers are trying to build more sophisticated models to represent the physical processes of soil particle detachment, transport, and deposition (Rose 1993).

An alternative, purely empirical approach relates changes in the sediment load of a river to changes in land cover in the surrounding watershed. The empirical approach is an essential check on theoretical models, but lack of data usually makes it hard to apply. One exception is a study by Alford (1992) that assembled annual time series data on sediment transport, streamflow, and precipitation for the Ping River in Northern Thailand from 1958 to 1985. Despite a decline in forest cover from 92 percent in 1973 to 73 percent in 1991 in Chiang Mai Province, sediment concentration in the Ping was approximately constant.[2] According to Alford, the near-linear relation between streamflow volume and total sediment transport "implies a sediment source within the stream channel rather than erosion from slopes

contributing sediment to this channel" (p. 267). This somewhat surprising conclusion—that significant deforestation was not accompanied by increased sedimentation—underscores the need for empirical studies to explore the role of local geology and topography in modulating the effects of changes in land use on sediment delivery.

Calculating the Extent of Economic Damage to Dams

The accumulation of river-borne sediment deposited in dams reduces the active storage volume of the reservoir and by so doing slows the output of irrigation, hydroelectric, and flood control services (Mahmood 1987; Southgate and Macke 1989). Moreover, sedimentation limits the effective life of the dam by advancing the date at which capacity is exhausted (Southgate and Macke 1989). Silt also damages turbines and increases the need for dredging. The total costs of siltation are significant; Mahmood (1987) puts annual global costs of lost reservoir capacity at about $6 billion. Chunhong (1995) reports that sedimentation reduces the storage capacity of China's reservoirs by 2.3 percent annually. The relevant question, though, is the marginal effect of deforestation-related sedimentation. The benefits of forest preservation in a dam's catchment area depend not only on the amount of sediment generated but also on the per hectare benefits provided by the dam, whether sediment is directed past the dam, and the timing of sediment-related damage (if not averted).

Consider two hydroelectric plants with the same generating capacity, one in a small steep-sided watershed, the second in a broad, shallow, gently sloping one. Forest protection will prevent erosion and sedimentation in both. The benefits of saving a forest hectare will be greater in the narrow watershed not only because erosion rates are higher on hills and sediment delivery faster in steep narrow valleys, but also because the dam benefits per hectare of watershed are higher.

• *Sediment management.* A variety of engineering and operational "fixes" can be used to sluice incoming sediments past dams or to flush accumulated sediments out of reservoirs (Lysne and others 1995; Chunhong 1995). These techniques are not universally applicable and typically have opportunity costs (in downtime or diminished output) in addition to capital costs. They also shift the costs of sedimentation downstream. To the extent that dams can minimize sedimentation using these techniques, however, the value of erosion prevention strategies such as forest preservation are reduced.

• *Time path of benefits.* Because siltation is a gradual process, measuring the amount of damage it does depends on how much time elapses before the buildup reduces the dam's effectiveness. Three potential time lags can occur between the initiation of a change in land use and a diminution in the benefits provided by the dam. First, the rate of land-use change matters. Clearcutting or the construction of low-quality logging roads could generate substantial amounts of erosion quickly. Conversely, in-

creasing the intensity of shifting cultivation might take decades to make a substantial change in basin-wide erosion.

Second, it takes time for sediment to travel. An eroded soil particle works its way down a watershed through a process of redeposition and resuspension. The amount of time between initial erosion and arrival in the reservoir depends on the gradient of the stream as well as on distance. Harden (1993) notes in connection with the Paute watershed in Ecuador that "sediment eroded from agricultural lands in distant, low-gradient tributary catchments may not reach the reservoir in the next half-century, but increased sediment loads in proximal, high-gradient tributary rivers represent an immediate sedimentation hazard" (p. 183). Conversely, sediment may continue to flow into rivers for twenty to thirty years after source erosion stops (Bruijnzeel 1990, Mahmood 1987).

The third lag is the time between the arrival of the sediment at a reservoir and the drop in dam output. Although reservoirs are built with dead storage capacity designed specifically to catch sediment, significant amounts are deposited in the active storage area, potentially reducing output. Additional sediment would thus be expected to affect dam services immediately. Southgate and Macke (1989) found, however, that earlier retirement of the dam (as opposed to decreased output before retirement) accounted for 85 percent of the economic impact of an increase in sedimentation rates.

Although these processes are quite complex, a simple numerical example illustrates the sensitivity of economic impacts to assumptions about the timing of deposits and discounting of dam benefits. Assume that dam services are constant until the dam is retired and that the effect of watershed damage reduces the expected lifetime of the dam from 100 years to 61 years. Given a 10 percent discount rate, the net present value of watershed protection is about 2.2 percent of the annual flow of dam benefits. A modest increase in the discount rate, to 12 percent, decreases the present value of watershed protection by 75 percent. The introduction of a twenty-year lag between the change in land use and the onset of sediment inflows decreases the present value of protection by a further 95 percent, to just 0.04 percent of the annual benefit flow.[3] In short, the benefits of extending the life of a relatively young dam will tend to occur in the distant future and therefore will be highly discounted.

Empirical Studies

The theoretical bent of this discussion on dams reflects a paucity of empirical studies. Only five estimates of the economic impact of changes in land use on dam performance can be found: Briones (1991); Cruz, Francisco, and Conway (1988); Southgate and Macke (1989); Veloz and others (1985); and Shahwahid and others (1997). The difficulty of gathering primary data on erosion processes is evident. Only the last study is based on empirical analyses relating actual sedimentation to

actual changes in land use. Only one study allows for a lag between the project's initiation and the impact of increased sediment. On the economic side, only one attempts to model in detail the process by which sediment reduces the life of a dam, and none incorporates the models used by dam engineers to describe the patterns of sediment buildup in reservoirs.

The per hectare benefits differ widely among these five studies. The highest value by far, more than $2,000, refers to a subset of the interventions envisioned for the Dominican Republic's Valdesia watershed management project, namely, the reforestation of the steepest slopes in the watershed. Some of the assumptions underlying this estimate are open to question, and the 5 percent discount rate elevates the value compared with some of the other studies. Nonetheless, this example illustrates the potential for very high levels of domestic benefits from protecting critical watershed areas.

In contrast, Cruz, Francisco, and Conway (1988) found that erosion around the Pantabangan dam in the Philippines resulted in loss of dam services equivalent to about $4 a year per hectare of converted forest. The implication is that forest protection would have provided benefits of that magnitude—somewhat less than $80 per hectare if capitalized at 5 percent.[3] These forest benefits, however, are not netted against the opportunity costs of restricting agricultural use, so net domestic benefits are lower. At the same time, Cruz, Francisco, and Conway argue that maintenance of forest cover would also have yielded substantial additional benefits by allowing the dead storage capacity of the dam to be converted to active storage for irrigation

The lowest net value for forest protection is reported by Shahwahid and others (1997) in an analysis of the Hulu Langat Forest Reserve in Malaysia. Here the alternative to complete forest protection is permitting low-impact logging of the forest (A noteworthy feature of such logging is that it prohibits cutting within twenty meters of a river or stream; according to the authors, this restriction reduces logging-related sedimentation by 60 percent, even though the stream buffers occupy less than 20 percent of the forest area.) In this case forest protection yielded a gross annual benefit of $44 per hectare. The opportunity cost of prohibiting logging was about $1,400 per hectare, however, reflecting the high density of commercial tree species in Malaysian forests. Thus the net benefits of forest protection, relative to low-impact logging, are –$1,356. Low-impact logging, as a land use, might have high benefits relative to alternative land uses, however; these have not been evaluated.

Impact on Fish and Aquatic Organisms

In an analysis of carefully gathered primary data, Hodgson and Dixon (1988) examined the impact of sedimentation on marine life in Palawan, the Philippines. They found that logging in the area had led to the construction of highly erosion-prone roads quite close to the Manlag River, within a few kilometers of Bacuit Bay. The

consequence was a very large and immediate increase in sediment that was "often more than 1,000 milligrams per liter," while sedimentation levels in a control river "rarely exceeded 10 milligrams per liter." The increased sedimentation destroyed nearly 50 percent of the coral cover on the reef nearest the mouth of the river. Although the levels of sediment were not high enough to kill the fish directly, coral mortality severely disrupts the ecosystems on which the fish depend.

Hodgson and Dixon impute the per hectare value of forest protection at a high $3,200, which overestimates the social gain because it is based on gross revenues from fisheries and tourism rather than on net profits. It is also worth noting that the authors rule out, as infeasible, interventions to reduce road-generated erosion, even though such interventions may save the loggers money by reducing maintenance costs. Because roads generate the bulk of all sediment, improved road-building techniques might make logging, fisheries, and tourism mutually compatible.

Erosion and Agricultural Productivity

If deforestation causes an increase in on-site erosion and a loss of agricultural productivity, can that loss be translated directly into forest preservation benefits? The answer is no, if forests and crops are mutually exclusive land uses. Once the forest has been converted to agriculture, erosion diminishes agricultural yields. But that rate of diminution is not the benefit of forest preservation.

Forest preservation, however, can yield agricultural productivity benefits through erosion reduction in two situations. First, some woodlands or open forests are used for grazing or cropping. Removing the trees in these areas to intensify production could be self-defeating if erosion increases drastically. Second, deforestation could result in increased runoff and thereby increase erosion on downslope croplands. This seems plausible and may be important in some areas, but we can find no relevant studies. Deforestation might also cause downslope damage from landslides. Conversely, erosion sometimes delivers valuable soils from uninhabited hillsides to farmers' fields (Enters 1992). Where that is true, forest preservation imposes external costs on those farmers. But these effects may be limited to exceptional soil conditions.

Impact of Land-Use Changes on Water Yield

Popular belief and casual empiricism link deforestation with flooding. If it is true that upslope deforestation threatens downstream cities and croplands with flood damage, the gains to forest preservation might be quite large. In fact, extensive scientific evidence links deforestation to annual increases in water yield (that is, the total volume of surface runoff and subsurface flows). As in the case of erosion, the increase depends on how the land is used (Bruijnzeel 1990, pp. 82–92). But increases in the

average rate of flow do not necessarily correspond to increases in peak flow or storm flow, which cause floods.

Surprisingly, the scientific literature supports a link between deforestation and flooding only at a local level—within a drainage basin of less than about 50,000 hectares (Bruijnzeel and Bremmer 1989). In small watersheds increases in water yield translate directly into increases in storm flow. For larger drainage basins, however, the limited number of studies using long time-series data on floods show no link between deforestation and flooding. Bruijnzeel (1990) cites studies of medium-size drainage basins (up to 1.45 million hectares) in Taiwan (China), and Thailand, which show that extensive deforestation had no effect on flooding, and three studies of India for the period 1871–1980, which show no trend in the frequency of flooding despite massive changes in land use during this time. Bruijnzeel and Bremmer (1989) also argue that there is no relation between changes in land use in the Himalayas and flooding in the Ganges-Brahmaputra basin, although they do not present time-series data. And Anderson, da Franca Ribeiro dos Santos, and Diaz (1993), who analyze eight decades of time-series data on rainfall and storm flow in the Parana-Paraguay river basin, report no structural shift in the relation between intense rainfall and floods, despite the significant conversion of forests over that period to pasture and cropland.

At first glance, these results seem paradoxical: How can deforestation cause flooding in small basins but not in large basins? The hypothesis is that basin-wide flooding depends more on intensity of the rainfall than on the way land is used. Most storms are small and transient. Individual subbasins tend to flood in sequence, as the storm passes over, rather than simultaneously. Local floods are thus averaged out over space and time. Only extremely severe and long-lasting storms affect all the tributaries of a major river at once. Storms of that magnitude would be large enough to saturate the soil's absorptive capacity and cause rapid runoff even if the land were still forested (Hamilton 1987; Bruijnzeel and Bremmer 1989; Bonell and Balek 1993, pp. 227–28).

Impact of Land-Use Change and Dry-Season Flows

Since the time of Plato, it has been assumed that deforestation results in lower water tables and reduced flows of water during the dry season (Grimble, Aglionby, and Quan 1994). This belief is still current; Huntoon (1992) links the loss of the "green reservoirs" of hillside forests in South China to severe reductions in the availability of groundwater during the dry season.

According to current hydrological science, however, the effects of deforestation on dry-season flows are ambiguous but likely to be counterintuitive. (Bonell and Balek 1993; Bruijnzeel 1990). This is because the conversion of forests to other purposes has two opposing effects on the water table. On the one hand, it increases

runoff and decreases the absorption of water into the ground. This by itself would lower the water table. On the other hand, trees are highly effective water pumps, removing water from the soil and transpiring it into the air. The replacement of trees by vegetation with shallow roots and lower transpiration rates (such as grass, annual crops, or early stages of secondary regrowth) therefore tends to reduce groundwater loss and raise the water table. Dozens of controlled experiments have been conducted showing that, contrary to expectations, the net immediate effect of tree removal is a rise in the water table, and therefore a probable increase in dry season flows (Hamilton and King 1983). Similar results have been found in studies of actual sites. Nepstad and Schwarzman (1992) compare deep-rooted evergreen forests to an adjacent degraded pasture in Pará, Amazonia. At the end of the dry season, the water in the top eight meters of soil available for plants was 370 millimeters higher in the degraded pasture.

In an interesting case study on Thailand, Vincent and others (1995) found that reforestation reduced, rather than increased, dry season flows and imposed costs on downstream users. Starting in 1967 Thai authorities promoted reforestation and sedentary agriculture in deforested areas of the Mae Theng watershed. These efforts involved two water-consuming interventions: the construction of irrigation systems and the establishment of pine plantations, which transpire more water than the deciduous forests that originally covered the area. An analysis of monthly stream flow records showed no change in dry season flows from 1952 to 1972 but registered a significant reduction from 1972 to 1991, when annual stream flows slowed by an additional 2.9 million cubic meters each year. These reductions resulted in the seasonal closure of one of Chiang Mai's water treatment plants and forced downstream farmers to switch from rice to soybean cultivation. The marginal costs of these reductions in water availability ranged from about 1 baht per cubic meter for agriculture to 7 baht for industrial users. These results imply that upslope deforestation, while highly undesirable on many grounds, yielded external benefits rather than costs for downstream water users.

Under some circumstances, however, deforestation may indeed reduce water tables. Bruijnzeel (1990) and Bonell and Balek (1993) point out that many processes involved in forest conversion compact the soil and cause gullying. Such processes include overgrazing, road construction, and the use of heavy machinery for land clearance. Compaction and gullying, in turn, increase runoff and decrease infiltration. If infiltration is reduced more than transpiration, the water table could drop.[5] Hamilton and King (1983) cite Australian studies showing severe reductions in infiltration following heavy grazing. They also cite a Fiji study finding runoff rates of 90 percent on grassland. They were unable to find analogous results, however, after forests were converted to annual cropping. A different situation is described by Kumari (1995). In this case, selective logging of a peat swamp forest entailed the construc-

tion of drainage canals. Expansion of the drainage network could reduce water storage sufficiently to imperil dry-season rice production in adjacent fields.

Climate Maintenance

There is a long-standing belief that deforestation reduces rainfall. Grove (1994) provides a fascinating account of scientific and policy interest in the topic dating to the seventeenth century (see, for example, Halley 1694). To the modern observer, too, it seems intuitively obvious that tropical deforestation reduces rainfall. Because evapotranspiration from tropical forests makes up between 20 percent (Southeast Asia) and 80 percent (Africa) of incident rainfall (Wilkie and Trexler 1993), it seems logical to expect that forest removal would break this recycling process, resulting in a drier climate.

Modern climate theory, however, introduces a host of additional complexities. Changes in land cover introduce not only changes in evapotranspiration, but also in albedo (surface reflectivity) and aerodynamic drag. These changes directly affect temperature and precipitation and also set off a whole round of positive and negative effects involving changes in cloudiness, air circulation patterns, and even plant transpiration. The result is a highly nonlinear, scale-dependent, dynamic system. No longer is it clear a priori that deforestation reduces local rainfall. Eltahir and Bras (1992) suggest, for example, that deforestation on the scale of hundreds of square kilometers increases convection and therefore rainfall, while deforestation on the scale of millions of square kilometers reduces rainfall. In any case the magnitude and spatial distribution of climate effects will be sensitive to local conditions, especially to the nature of the vegetation that replaces the forest.

Theoretical analysis of the climatic impact of changes in land use therefore requires sophisticated models. During the past ten years, general circulation models of the earth's atmosphere have been used to analyze the effect of large-scale deforestation on global climate. Several exercises have examined the implications of converting the entire Amazon or Southeast Asian rainforests to savanna. In principle, these exercises might be used to evaluate the domestic benefits of forest preservation for large countries such as Brazil or Indonesia. Henderson-Sellers and others (1993) predict that complete deforestation of the Amazon would reduce precipitation in the rainy season by 30 percent, while complete deforestation of Southeast Asia would have no effect on precipitation. Lean and Rowntree (1993) predict that total deforestation of the Amazon would reduce local rainfall by 14 percent, but increase rainfall in Eastern Brazil by 20 percent.

These results must be interpreted with caution for several reasons. First, the scale and permanence of the deforestation simulated in these exercises is unrealistic. Shukla, Nobre, and Sellers (1990), for instance, assume that the entire Amazon would be reduced to degraded pasture. Many of the deforested areas in the Amazon in fact

revert to secondary forest (Moran, Mausel, and Wu 1994), whose climatological properties are much closer to primary forest than to pasture. Second, despite their sophistication, general circulation models omit a range of physical processes and rely on a great many assumptions about parameters. How sensitive the results are to these omissions and assumptions is unknown, however. Third, these models divide the planet's surface into a very coarse grid and can only be applied to deforestation processes at the scale of tens of thousands—or more—of square kilometers. Results at this scale cannot be generalized to deforestation patches of tens or hundreds of square kilometers. Work on more appropriate mesoscale models is still in its infancy.

Empirical work in this area is as inconclusive as the theoretical work. Bruijnzeel (1990) reviews the thin literature. There are a great many microstudies of temperature and soil wetness changes in small clearings, but these cannot be generalized to larger scales and are useful mainly to establish the parameters for general circulation and mesoscale models. A limited number of mesoscale empirical studies try to relate changes in forest cover to changes in recorded precipitation (see, for example, Meher-Homji 1988), but Bruijnzeel notes that these are lacking in rigor and data consistency.

Current research is just beginning to use remote sensing data to track climatic and land-use changes, resulting in more rigorous studies. Complementing twenty-two years of data from twenty climate stations in the Selva Lancondona region of Chiapas, Mexico, O'Brien (1995) used remote sensing data from 1979 and 1989 to track deforestation around each station. Preliminary analysis indicates that deforestation increases minimum temperature, decreases maximum temperature, and has no significant effect on precipitation. Cutrim, Martin, and Rabin (1995) use satellite data to show increases in cloudiness (not necessarily implying increased precipitation) following large-scale deforestation in Rondonia, Brazil.

In sum, the assumption that deforestation affects local climate is plausible, but the magnitude (and indeed sign) of the effect remains to be demonstrated. The potential economic significance of climatic effects should make them an early priority for research.

Commercial Value of Nontimber Forest Products

The value of tropical nontimber forest products, such as fruits, nuts, latex, resins, medicines, and animals, has been reviewed by Godoy, Lubowski, and Markandya (1993) and Lampietti and Dixon (1995). In a survey of twenty-four studies, Godoy, Lubowski, and Markandya report per hectare values ranging from \$.75 to \$422 a year, with a median of about \$50. But these studies, as a group, exaggerate the level of benefits that would accrue to the domestic economy from a typical natural, or old growth, forest. They do not distinguish between agroforestry and pure extraction, nor do they take into account the costs of extraction. Furthermore, they do not allow

for spatial differences in the densities and therefore quantities of the product or in the rate of extraction. And they fail to allow for competition in the markets for these products in the long term.

The Distinction between Agroforestry and Extraction

Nontimber forest products can be produced at different levels of intensity, with correspondingly different degrees of disturbance of the original ecosystem. At one extreme are purely extractive systems, where extractors harvest products from an otherwise undisturbed forest. In a slightly more intensive approach, extractors may artificially enrich the forest with a desired plant species. Still more intensive are a range of agroforestry techniques that replace the primary forest with carefully manipulated multispecies plantations to provide raw materials for trade and industry.

To give a true picture of the benefits of preserving "natural" forest, nontimber forest product valuation must be based on the profits from purely extractive systems, rather than profits from agroforestry systems. It is tempting to rely on the latter. Agroforestry systems typically generate higher values per hectare because commercial species are planted more densely and extraction and processing costs are lower. But although these systems are attractive for many reasons, including their relatively high degree of biodiversity, they are not the same as the natural ecosystems they replace and therefore should not be used to justify preservation of the natural forest.

At the same time, it is worth stressing that agroforestry systems may offer both greater biodiversity and higher economic benefits than other uses of the land. This is true of a Sumatran system in which slash-and-burn farmers create rubber-rich secondary forests. After the rubber trees reach maturity (about ten years), they yield about 600 kilograms per hectare annually of dry-equivalent rubber (van Noordwijk and others 1995, p. 88) with no inputs other than labor (current rubber prices are about $1.60 a kilogram). Extraction can continue for twenty years or more before another cycle of clearing. Because the owners' share in the typical tapping arrangement is one-third, per hectare rents are substantial. At the same time, ecological studies show fairly high levels of species richness (Michon and de Foresta 1995; Thiollay 1995). Moreover, these rubber forests should have the same hydrological properties as primary forests. Thus agroforests may in some cases offer both net domestic economic benefits and global or nonmarket benefits relative to competing land uses.

Allowing for the Costs of Extraction

Conceptually, the value of a hectare of forest for nontimber products is equivalent to the rent that would be paid for the right to harvest that hectare. Clearly that amount is less than the final value of the product in the marketplace because the costs of

extraction and transportation must be netted from the sales price to yield profit or rent. This elementary point has been made many times in the literature and must be applied to the cases presented by Godoy, Lubowski, and Markandya (1993). For instance, they find the highest documented value of a hectare of forest based on actual extraction rates is Chopra's (1993) estimate of $117–$144 a year for fuelwood, fodder, and miscellaneous products from tropical deciduous forests in India. In the absence of market prices for these goods, Chopra values them either by their cost of extraction or by the price of substitute commodities. For instance, he values labor to gather fuelwood at $18.87 to $24.17 a hectare (while its equivalent in softcoke, an alternative fuel, would cost $9.50 to $17.33). Chopra concludes that the value of forests for fuelwood must lie between $9.50 and $24.17 a hectare. But this conclusion confuses costs, benefits, and rents. If softcoke is in fact a close substitute for fuelwood, then the data imply that villagers are expending labor worth more than $18.87 to produce fuelwood worth less than $17.33. Clearly the estimates are crude, but the main implication is that the net per hectare value of the forest for firewood production is close to zero. Similarly, labor expenditures to produce miscellaneous goods such as lacquer and dyes amount to $66.67 a hectare. To calculate the value of the forest in producing these goods, it is thus necessary to subtract $66.67 a hectare from the price paid for the lacquer and dyes.

Spatial Variation in Density and Extraction Cost

Forests tend to be large and heterogeneous. Estimating the value per hectare for a small plot and attributing that value to the forest as a whole (let alone to any other forest) is inappropriate. This point is obvious but is almost universally ignored in practice.

Three difficulties arise in generalizing small-plot estimates. First, transport costs are important for some nontimber forest products; açai fruit, for example, spoils within twenty-four hours of harvesting. The high costs of transporting bulky or perishable goods through the forest means that the value at the point of collection will fall off steeply with distance from the road and from the market. Large portions of the forest will not be economically exploitable for this class of commodities and will thus have zero value for that use. At the same time high-value, less perishable commodities, such as rubber, will be economically viable over a larger area.

Second, the density of exploitable species can vary dramatically within and between forests. For instance, two of the highest per hectare values cited in Godoy, Lubowski, and Markandya (1993) refer to oligarchic forests, that is, those dominated by a few, highly commercial species (Anderson and Ioris 1992; Anderson and Jardim 1989). Although interesting and locally important, such examples are quite atypical of tropical rainforests, whose hallmark is very high diversity and thus very low densities for any individual species.

Third, consumption of some nontimber forest products may tap only a sm fraction of the potential supplying area. In Ecuador, for instance, Grimes and oth (1994) studied a resin derived from a *Protium* tree. Harvesting the trees yielded average potential net return per hectare (after collection, transport, and marketin of $61 a year in the three forest plots surveyed. The resin, however, is used excl sively for finishing local ceramic handicrafts, which are presumably in limited d mand. We surmise that the total number of hectares being harvested is a small fra tion of the total area from which the trees could be economically harvested.[6] If so, would be a gross error to apply the $61 value to the entire range of the species. Tl same situation may apply to valuable medicinal plants.

Long-Run Competitive Supply

For most commercially attractive products, pure extractive reserves cannot compe with synthetic or domesticated substitutes (Richards 1993; Browder 1989). Ma nontimber forest products follow a life cycle in which they start out as extracti products, attain a world market and a high price, and are then domesticated in i tensive plantation or agroforestry systems. Intensive cultivation reduces labor, lan and capital costs; permits product standardization; facilitates processing; ensures reliable supply; and takes advantage of scale economies in marketing. As a result, tl supply price of the product falls below the viable level for extractive supply. Tl prime example is rubber. In the early part of this century, Brazilian rubber pri (and extraction volumes) collapsed as Malaysian plantations came on line. In rece years Brazil's extractive reserves have been supported by subsidies. Currentl extractivists in that country's Chico Mendes reserve are being driven out of busine by lower-priced latex produced by plantations in Sao Paulo State. Labor producti ity in the plantations is about ten times that in the reserve (Brooke 1995).

The implication is that lower-cost domesticated or synthetic substitutes grea reduce—or even eliminate—the rents from extractive reserves. Nontimber fore products provide domestic benefits only when the products are difficult to domest cate or duplicate.

This gloomy statement requires some qualification, however. First, those forest in which a few commercially valuable species grow may be able to compete wit plantations, especially with some small interventions such as pruning (Anderso and Jardim 1989). Second, agroforestry systems such as jungle rubber may in som cases be competitive with plantations while also preserving some biodiversity. Thir and most important, the success of large-scale plantations for nontimber product increases the global urgency of preserving genetic diversity. This may often be be accomplished through preserving natural forests. The implication is that th global benefits of forest preservation may increase even as the domestic benefit decline.

The Opportunity Costs of Preservation

The sustainable benefits associated with forest preservation can be thought of as the gross benefits. The existence of a threat to a forest tract usually implies an economic motive for converting the forest or exploiting it, and these potential benefits are the opportunity costs of preservation. These opportunity costs must be deducted from gross forest benefits to yield the net domestic benefits of preservation.

Opportunity costs are highly sensitive to land characteristics. The returns to agricultural use of a plot depend on the physical characteristics of the land, current vegetation, market access, and land tenure. Physical characteristics such as slope, drainage, and soil fertility determine the land's relative physical productivity for different crops, the need for inputs, and the degree to which output can be sustained over time. The density of commercial tree species also affects the net cost of clearing; in some cases the value of timber may outweigh the benefits from agriculture. Market access — the cost of transport to the nearest market — determines the potential price paid to the farmer for crops or cattle and their inputs. Land tenure and ownership, together with land-related tax and subsidy rules, affect the incentives to invest in land preparation and in perennial crops.

The opportunity cost of land—that is, the forgone benefits of conversion—can be obtained through Geographic Information Systems data and techniques. Using such data Chomitz and Gray (1996), for instance, found that the type of soil and the accessibility to roads strongly influence the probability that a forest will be converted. Magrath and others (1995) used farm budgets and a land capabilities assessment to impute land values in West Kalimantan, Indonesia. The land capabilities assessment divided the province into 1,682 polygons and assessed the suitability of each polygon for a variety of crops. (For lack of data, however, no adjustments were made for the cost of transporting crops to markets or for benefits from timber marketing.) The results are quite striking: much of West Kalimantan's land has little value for agricultural production. Of the province's 14.65 million hectares, 3.7 million have an opportunity cost of less than 20 cents a hectare a year (1991 prices). About 95 percent of the province has an agricultural opportunity cost of less than $2 a hectare annually. Were transport costs factored in, the opportunity costs would be still lower.

Summary of Findings

The level of net domestic benefits from forest preservation depends on the alternative land use as well as local climatic, biological, geological, and economic circumstances. When the alternative use is agroforestry or forest plantations (depending on

the management system), preservation of the natural forest may not offer net benefits from hydrological benefits or the production of minor forest products. At the same time, some agroforests may offer both biodiversity benefits and net domestic economic benefits relative to other land uses.

The prospects for economically significant hydrological benefits from forest preservation appear to be smaller than popularly supposed.

- Deforestation has not been shown to be associated with large-scale flooding, although it may cause serious local flood damage.
- Nor is it associated with diminished dry season flows; on the contrary, it is usually associated with greater flows.
- Although it is a priori plausible that deforestation should affect local precipitation, the magnitude and even the direction of the effects are not known except in the special case of cloud forests, which "harvest" passing moisture.
- The link between deforestation and downstream sediment damage is sensitive to the basin topography and geology. Where sediment transport is slow—as in large, low-gradient basins—downstream effects may occur far in the future, so that the net present value of damages is small.

Conversely, forest preservation can yield substantial domestic benefits where it averts erosion-generating changes such as road building, annual cropping, or overgrazing; where affected areas impinge directly on streams, reservoirs, coral reefs, or inhabited areas; and where affected watersheds are small, steep, and erosion prone.

The measurement of benefits from harvesting minor forest products is still rudimentary. Current valuation exercises tend to be specific to particular plots and cannot be generalized to significant forest areas for one of several reasons: a) they are based on inventories of salable products rather than actual extraction; b) their value is based on gross market prices rather than prices net of extraction and transport costs; c) the study describes agroforestry products rather than true extractive products; d) the study describes products of unusual forests dominated by a few commercial species rather than a typical species-rich rain forest; and e) the value of the forest for extractive production of commercial nontimber forest products is undercut by competition from domesticated or synthetic substitutes.

We stress again that our review of benefits is not comprehensive. Potential domestic benefits not reviewed here include ecotourism services, sales of bioprospecting rights, and carbon sequestration services, should a carbon offsets market come into existence.[7] Moreover forest preservation can yield substantial global or ecological benefits. Also note that these conclusions apply only to tropical moist forests. It may be the case that for other ecosystems, such as wetlands, the links between land use change and economic benefits are both better understood and stronger.

Domestic economic benefits provide an uncertain rationale for conservation—and especially for funding forest preservation through market-rate loans. Undoubt

edly this rationale is clearly justified for some forest preservation projects, and these should be vigorously pursued. For many projects, however, net domestic benefits either do not exist or cannot be quantified with sufficient rigor to support a market-rate loan or a convincing cost-benefit analysis. Hence the domestic benefits argument—save your forests because they bring palpable economic benefits to your country—cannot be the mainstay of forest preservation in all countries. For many—possibly most—tropical forests, the more compelling rationale for preservation is based on global values. This underscores the need for financing conservation from the Global Environment Facility or other global sources, rather than placing the burden entirely on domestic resources.

The hopeful converse, however, is that the net domestic costs of forest preservation may also be small. The argument is: Save your forests because the out-of-pocket costs of doing so are small, and the noneconomic benefits are large. Many biodiverse, carbon-rich forest areas are poorly suited to agriculture because of isolation and poor soils. These areas can be preserved through a three-pronged strategy. First, the opportunity costs of preservation should be kept low by directing regional development toward more economically promising districts. Above all, uneconomic road-building should be avoided in these areas. Once roads are in place, the opportunity costs of preservation can increase substantially. Second, where pressures for logging are politically and economically irresistible, low-impact techniques can be required as a condition for access. This approach would entail disabling main access roads after logging was completed. Third, direct and ongoing compensation can be paid, or alternative livelihoods set up, for land users or stakeholders who would otherwise convert the forest to other uses.

Notes

Kenneth M. Chomitz is with the Development Research Group of the World Bank, and Kanta Kumari is with the Secretariat of the Global Environmental Facility. Preparation of this article was supported by the World Bank's Policy Research and Environment Departments. The authors are grateful to Thomas Walters, William Hyde, Nalin Kishor, and Lini Wollenberg for helpful comments.

1. Induced sedimentation may be a concern even when background sedimentation levels are higher if marginal increases in sediment result in further damage and if the costs of averting such sedimentation are smaller than those of remedying background sedimentation.

2. Forest cover data supplied by Charles Griffiths, U.S. Environmental Protection Agency.

3. The assumption that dam life with lagged sediment delivery is eighty-one years is very crude but will suffice for illustrative purposes. Also, note that a ratio of 0.04 percent does not necessarily imply that a watershed protection project is uneconomic—the project's value depends on the cost of protecting the watershed, which might be relatively small.

4. Authors' calculations based on Cruz, Francisco, and Conway data.

5. This does not explain Huntoon's reports for South China, because the deforestation resulting from the felling of trees would not be expected to result in soil compaction.

6. What then sustains the net return of $61 a hectare? Why doesn't competition drive these returns toward zero? Three answers are possible: the producing plots are to some extent, perhaps

informally, privatized, and the return reflects the opportunity cost of the land; the $61 figure includes returns to the expertise of the collector, who knows how and where to find it; or in fact all economically exploitable areas are being harvested, and the $61 measurement is based on inframarginal plots.

7. But see Simpson, Sedjo, and Reid (1996) for a pessimistic view.

References

The word "processed" describes informally reproduced works that may not be commonly available though library systems.

Alcorn, J. B. 1989. "An Economic Analysis of Huastec Mayan Agroforestry Management." In Browder (1989).

Alford, Donald. 1992. "Streamflow and Sediment Transport from Mountain Watersheds of the Chao Phraya Basin, Northern Thailand: A Reconnaissance Study." *Mountain Research and Development* 12(3, August):257–68.

Anderson, A. B., and E. M. Ioris. 1992. "The Logic of Extraction: Resource Management and Income Generation by Extractive Producers in the Amazon Estuary." In K. H. Redford and C. Padoch, eds., *Conservation of Neotropical Forests. Working from Traditional Resource Use.* New York: Columbia University Press.

Anderson, A. B., and M. A. G. Jardim. 1989. "Costs and Benefits of Floodplain Forest Management by Rural Inhabitants in the Amazon Estuary: A Case Study of Açaí Palm Production." In Browder (1989).

Anderson, Robert J., Jr., Nelson da Franca Ribeiro dos Santos, and Henry F. Diaz. 1993. "An Analysis of Flooding in the Parana/Paraguay River Basin." Dissemination note 5. World Bank Latin America and the Caribbean Technical Department, Washington, D.C. Processed.

Bonell, Michael, and J. Balek. 1993. "Recent Scientific Developments and Research Needs in Hydrological Processes of the Humid Tropics." In Bonell, Hufschmidt, and Gladwell (1993).

Bonell, Michael, Maynard M. Hufschmidt, and John S. Gladwell, eds. 1993. *Hydrology and Water Management in the Humid Tropics. Hydrological Research Issues and Strategies for Water Management.* Cambridge, U.K.: Cambridge University Press.

Botkin, Daniel B., and Lee M. Talbot. 1992. "Biological Diversity and Forests." In Narendra P. Sharma, ed., *Managing the World's Forests. Looking for Balance between Conservation and Development.* Dubuque, Iowa: Kendall/Hunt Publishing Co.

Briones, N. D. 1991. "Estimating Erosion Costs: A Philippine Case Study in the Lower Agno River Watershed." In Easter, K. W., J. A. Dixon, and M. M. Hufschmidt, eds., *Watershed Resource Management: Studies from Asia and the Pacific.* Environment Development Series. Singapore: Institute of Southeast Asian Studies.

Brooke, James. 1995. "Rubber Trees Being Planted in Brazil, the Source." *The New York Times* Jan. 9, p. 14.

Browder, John O., ed. 1989. *Fragile Lands in Latin America. The Search for Sustainable Uses.* Boulder, Colo.: Westview Press.

Bruijnzeel, L. A. 1989. "Deforestation and Dry Season Flow in the Tropics: A Closer Look." *Journal of Tropical Forest Science* 1:229–43.

__________. 1990. *Hydrology of Moist Tropical Forests and Effects of Conversion: A State of the Knowledge Review.* Amsterdam: Netherland Committee for the International Hydrological Programme of UNESCO.

Bruijnzeel, L. A., with C. N. Bremmer. 1989. *Highland-Lowland Interactions in the Ganges Brahmaputra River Basin: A Review of Published Literature.* Occasional Paper 11. Kathmandu, Nepal: International Center for Integrated Mountain Development.

Chomitz, Kenneth M., and David A. Gray. 1996. "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize." *The World Bank Economic Review* 10(3):487–512.

Chopra, Kanchan. 1993. "The Value of Nontimber Forest Products: An Estimation for Tropical Deciduous Forests in India." *Economic Botany* 47(3, July–September):251–58.

Chunhong, Hu. 1995. "Controlling Reservoir Sedimentation in China." *International Journal on Hydropower and Dams* 2(March):50–2.

Cruz, Wilfrido, H. A. Francisco, and Z. T. Conway. 1988. "The On-Site and Downstream Costs of Erosion in the Magat and Pantabangan Watershed." *Journal of Philippines Development* 26(25,1):85–111.

Cutrim, Elen, David W. Martin, and Robert Rabin. 1995. "Enhancement of Cumulus Clouds over Deforested Lands in Amazonia." *Bulletin of the American Meteorological Society* 76(October 1):1801.

Eltahir, E. A. B., and R. L. Bras. 1992. "On the Response of the Tropical Atmosphere to Large-Scale Deforestation." *Quarterly Journal of the Royal Meteorological Society* 119:779–93.

Enters, Thomas. 1992. "Land Degradation and Resource Conservation in the Highlands of Northern Thailand. The Limits to Economic Evaluations." Ph.D. dissertation. Australian National University, Canberra. Processed.

Godoy, Ricardo, Ruben Lubowski, and Anil Markandya. 1993. "A Method for the Economic Valuation of Nontimber Tropical Forest Products." *Economic Botany* 47(3):220–33.

Grimble, R. J., J. Aglionby, and J. Quan. 1994. *Tree Resources and Environmental Policy: A Stakeholder Approach.* NRI Socioeconomic Series 7. Chatham, U.K.: Natural Resources Institute.

Grimes, Alicia, and others. 1994. "Valuing the Rain Forest: The Economic Value of Nontimber Forest Products in Ecuador." *Ambio* 23(7):405–10.

Grove, Richard H. 1994. "A Historical Review of Early Institutional and Conservationist Responses to Fears of Artificially Induced Global Climate Change: The Deforestation-Desiccation Discourse 1500–1860." *Chemosphere* 29(5):1001–13.

Halley, Edmund. 1694. "An Account of the Watery Circulation of the Sea, and of the Cause of Springs." *Philosophical Transactions of the Royal Society* 142(7):468–72.

Hamilton, L. S. 1987. "What Are the Impacts of Himalayan Deforestation on the Ganges-Brahmaputra Lowlands and Delta? Relations between Assumptions and Facts." *Mountain Research and Development* 7:256–63.

Hamilton, L. S., and P. N. King. 1983. *Tropical Forested Watersheds: Hydrologic and Soils Response to Major Uses or Conversions.* Boulder, Colo.: Westview Press.

Harden, C. P. 1993. "Land Use, Soil Erosion, and Reservoir Sedimentation in an Andean Drainage Basin in Ecuador." *Mountain Research and Development* 13(2):177–84.

Henderson-Sellers, A., R. E. Dickinson, T. B. Durbidge, P. J. Kennedy, K. McGuffie, and A. J. Pitman. 1993. "Tropical Deforestation: Modeling Local to Regional-Scale Climate Change." *Journal of Geophysical Research* 98:7289–315.

Hodgson, Gregor, and John A. Dixon. 1988. *Logging versus Fisheries and Tourism in Palawan: An Environmental and Economic Analysis.* East-West Environment and Policy Institute Occasional Paper 7. Honolulu: East-West Center.

Huntoon, Peter W. 1992. "Hydrogeologic Characteristics and Deforestation of the Stone Forest Karst Aquifers of South China." *Ground Water* 30(2, March–April):167–76.

Kumari, Kanta. 1995. "An Environmental and Economic Assessment of Forest Management Options: A Case Study in Malaysia." Environment Department Paper 026. World Bank, Washington, D.C. Processed.

Lampietti, Julian A., and John A. Dixon. 1995. "To See the Forest for the Trees: A Guide to Nontimber Forest Benefits." Environment Department Paper 013. World Bank, Washington, D.C. Processed

Lean, J., and P. R. Rowntree. 1993. "A GCM Simulation of the Impact of Amazonian Deforestation on Climate Using an Improved Canopy Representation." *Quarterly Journal of the Royal Meteorological Society* 119(April 1):509–30.

Lysne, D. K., N. R. B. Olsen, H. Støle, and T. Jacobsen. 1995. "Sediment Control: Recent Developments for Headworks." *International Journal on Hydropower and Dams* 2(March 1995):46–49.

Magrath, W. B., Charles Peters, Nalin Kishor, and Puneet Kishor. 1995. "The Economic Supply of Biodiversity in West Kalimantan: Preliminary Results." World Bank, Asia Technical Department, Washington, D.C. Processed.

Mahmood, K. 1987. *Reservoir Sedimentation: Impact, Extent, and Mitigation* World Bank Technical Paper 71. Washington, D.C : World Bank.

Meher-Homji, V. M. 1988. "Effects of Forests on Precipitation in India." In Evan R C. Reynolds and Frank B. Thompson, eds., *Forests, Climate, and Hydrology: Regional Impacts.* Tokyo: United Nations University.

Michon, Geneviève, and Hubert de Foresta. 1995. "The Indonesian Agroforest Model: Forest Resource Management and Biodiversity Conservation." In P. Halladay, and D A. Gilmour, eds *Conserving Biodiversity outside Protected Areas: The Role of Traditional Agro-ecosystems* Gland Switzerland: IUCN, and Cambridge-UK.

Moran, Emilio F., Eduardo Brondizio, Paul Mausel, and You Wu. 1994. "Integrating Amazonian Vegetation, Land-Use, and Satellite Data." *BioScience* 44(5):329–38

Myers, Norman. 1995. "The World's Forests: Need for a Policy Appraisal " *Science* 268(May 12):82–83.

Nepstad, Daniel C., and Stephan Schwartzman, eds. 1992. *Nontimber Products from Tropical Forests Evaluation of a Conservation and Development Strategy. Advances in Economic Botany* Vol. 9 New York: The New York Botanical Garden.

O'Brien, Karen L. 1995. "Deforestation and Climate Change in the Selva Lacandona of Chiapas Mexico." *Norsk Geografisk Tidsskrift* 49(3):105–22.

Richards, E. M. 1993. *Commercialization of Nontimber Forest Products in Amazonia* NRI Socioeconomic series 2. Chatham, U.K.: Natural Resources Institute

Rose, C. W. 1993. "Erosion and Sedimentation." In Bonell, Hufschmidt, and Gladwell 1993

Shahwahid, H. O., and others. 1997. "Economic Benefits of Watershed Protection and the Trade-off with Timber Production: A Case Study in Malaysia." Research Report Series Economy and Environment Program for Southeast Asia, Singapore. Processed.

Shukla, J., C. Nobre, and P. Sellers. 1990. "Amazon Deforestation and Climate Change " *Science* 247(March 16):1322–25.

Simpson, R. David, Roger A. Sedjo, and John W. Reid. 1996. "Valuing Biodiversity for Use in Pharmaceutical Research." *Journal of Political Economy* 104:163–85.

Southgate, Douglas, and Robert Macke. 1989. "The Downstream Benefits of Soil Conservation in Third World Hydroelectric Watersheds." *Land Economics* 65(1):38–48.

Thiollay, Jean-Marc. 1995. "The Role of Traditional Agroforests in the Conservation of Rain Forest Bird Diversity in Sumatra." *Conservation Biology* 9(2):335–53.

van Noordwijk, Meine, Thomas P. Tomich, and others, eds. 1995. ***Alternatives to Slash-and-Burn in Indonesia: Summary Report of Phase I.*** ASB-Indonesia Report 4. Bogor, Indonesia: International Centre for Research in Agroforestry—Southeast Asia.

Veloz, Alberio, Douglas D. Southgate, F. J. Hitzhusen, and R. D. Macgregor. 1985. "The Economics of Erosion Control in a Subtropical Watershed: A Dominican Case." *Land Economics* 61:145–55.

Vincent, Jeffrey R., Mingsarn Kaosa-ard, and others. 1995. *The Economics of Watershed Protection: A Case Study of the Mae Teng River* Bangkok: Thailand Development Research Institute.

Wiersum, K. F. 1984. "Surface Erosion under Various Tropical Agroforestry Systems." In C. L. O'Loughlin and A. J. Pearce, eds., *Proceedings, Symposium on Effects of Forest Land Use on Erosion and Slope Stability* International Union of Forestry Research Organization, Vienna, and East-West Centre, Hawaii.

Wilkie, David S., and Mark C. Trexler. 1993. "Biogeophysical Setting and Global Climate Change." In *Central Africa: Global Climate Change and Development* Technical Report. Washington, D.C.: Biodiversity Support Program (a consortium of World Wildlife Fund, The Nature Conservancy, and World Resources Institute)

Poverty in Russia during the Transition: An Overview

Jeni Klugman • Jeanine Braithwaite

It is no surprise that the breakup of the Soviet Union and the overall demise of the planned economy has had a profound effect on the welfare of the Russian people. But the absence of reliable survey data has constrained our understanding of the impact that transition has had on the distribution of income. This article draws upon several rounds of a nationally representative household survey to document the sharp increases in the incidence and severity of poverty that have occurred during the transition. We investigate the routes by which macroeconomic and structural developments have been transmitted through the labor market and examine the performance of an increasingly overburdened, unfocused, and inadequate system of social protection. There is no evidence to suggest that the poor have shared in Russia's emerging economic recovery, and the emergence of a core group of long-term poor appears to be a distinct possibility.

The social and economic changes that swept across the former Soviet Union in the early 1990s were unprecedented in scale, scope, and speed. The collapse of an empire that spanned eleven time zones and encompassed more than 250 million people of a multitude of ethnic groups has had far-reaching consequences, both domestically and internationally.

The Russian Federation, which covers a vast expanse of Europe and Asia and includes a wide variety of social and cultural groups, naturally looms large in perceptions about the transition from a centrally planned to a market-oriented economic system. For many Russians that transition has been marked by the polarization of a previously egalitarian society and a dramatic increase in the scale of poverty and deprivation. The people who have been most adversely affected by structural changes in the economy were not necessarily so badly off under the old system, suggesting that some conventional assumptions about the structure of poverty need to be adjusted.

Underlying the sharp increases in the incidence and composition of poverty have been the overall decline in national income and the simultaneous increase in in-

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 37–58

equality. Aggregate shocks to output and productivity have been unevenly distributed among different groups in the economy. We trace the underlying mechanisms that reflect a shift from a system with controlled wages and incentives to maximize employment to a far less constrained environment, albeit at temporarily much lower overall levels of output and productivity. Adjustment in the labor market has affected overall earnings and wage disparities at the same time that the number of employed has declined along with total number of hours worked. Although the government social protection system has yet to adapt to the new economic conditions the fiscal constraints have hindered its ability to offset the increase in poverty. Growing regional disparities associated with economic liberalization have similarly strained the fiscal capacities of local authorities to effect redistribution.

As the economy stabilizes and clear signs of economic recovery emerge, the situation of the poor could be expected to stabilize and improve. Recent evidence suggests that the worst may be over, although the number of households and individuals below the poverty line continued to rise somewhat in 1995 and 1996. Beyond the direct recessional repercussions of the transition in Russia, the shift toward a market economy will have long-term implications for the poor. Changes in relative prices mean that certain groups (characterized, for example, by skill or location) are likely to be relatively better or worse off than they were before the transition. Inequality is also likely to remain higher than it was in the Soviet period, although the precise pattern of distribution will clearly depend on the government's tax and transfer policies.

This article first looks at the macroeconomic trends relevant to household welfare that have characterized Russia's transition. It then reviews the data sources and methodology—a critical dimension given the paucity of reliable data—and presents the major findings that have emerged from a recent analysis of poverty. After examining developments in the labor market and the limited effectiveness of public income transfers in alleviating poverty, we present some conclusions about social protection policy and its effect on poverty in Russia.

The Economic Backdrop

Russia inherited a command economy whose weaknesses had become manifest in the 1980s (Easterly and Fischer 1995). The economic story since 1991 is by now familiar to most readers. As the old political regime crumbled, inflation surged and output fell precipitously, leading to the collapse of internal and external trade and a rapid rise in the fiscal deficit. High inflation, which continued to 1995, severely depressed economic activity throughout much of the economy, and real earnings declined. Rapid consumer price inflation followed market liberalization in early 1992 and had adverse effects both generally, in accentuating the depth and duration of the

output recession (Popov 1996), and specifically, among those groups whose wages and transfer payments began to lag behind prices.

With the collapse of output, gross domestic product (GDP) fell more than 40 percent during 1991–96 (EBRD 1997). That decline places Russia among the worst of the formerly planned economies; estimated GDP decline ranged from 11 percent in Poland to 57 percent in Lithuania (de Melo, Denizer, and Gelb 1995). These estimates for Russia are problematic, however, because the structure of the decline in output and other measurement problems, as well as the failure to account for output from most informal sector activity, understate post-1991 output. Consistent with the pattern in other transition economies, there has been a relative expansion of the service sector's share of GDP (from 31 to more than 50 percent during 1989–94), especially trade and financial services, and a reduction in the share of investment and military production; part of this 'growth' in services, however, is attributable to changes in classification as well as in relative prices. Unofficial economic activity of both new private firms and privatized state enterprises has been expanding at a robust pace (Gavrilenko and Koen 1994). Industrial restructuring has been associated with a marked shift from secondary manufacturing into raw materials industries. The share of resource industries (energy, steel, and nonferrous metals) in industrial output doubled to 48 percent during 1991–95, while that of secondary manufacturing—particularly machinery, equipment, and light industries—declined.

These shifts have been associated with significant changes in the regional and interpersonal distribution of earnings and incomes. The regional concentration of production under central planning magnifies these subsequent distributive effects. Evidence suggests that the regional variation in nominal wages and incomes fell steadily during the two decades preceding the transition in Russia, but that regional disparities have increased sharply since 1991. The trend is even sharper when measured in real terms that account for significant geographical price variation (Stewart 1996). Indeed by 1994 income disparities among the oblasts (a political subdivision) of Russia were far greater than those among the states of the United States. The coefficients of variation for income among Russian oblasts and American states were 0.519 and 0.148, respectively. (The coefficient of variation shows the standard deviation divided by the mean. Values closer to one are less equal than values closer to zero.)

Measuring the social impact of the transition from a command economy is difficult, not least because the official economic data are unreliable (Koen 1996). Traditional sources of information on production, incomes, and prices, never totally reliable, have grown weaker, partly because many enterprises now fall outside the reporting network; official GDP figures were revised significantly every year between 1991 and 1994. Even more worrying has been the apparent divergence of time-series data that would normally be positively correlated (such as average money income and GDP). Thus the macroeconomic statistics cited here should be regarded with caution, and

the revealed trends regarded as indicative. Official poverty and income distribution measurements raise their own set of data problems.

Measuring Poverty in Russia: Conceptual and Empirical Issues

Assessments of household welfare confront a range of methodological and practical problems, from devising an appropriate definition of poverty to determining the appropriate measure of household welfare (using income, expenditure, or some non monetary indicator).

What Is Poverty?

Conceptually, few would dispute that poverty is the inability to sustain some minimal level of existence. The standard approach to defining an absolute level of poverty is to price a basket of essential goods and compare an individual's income to the cost of these necessities. But both in the definition of nutritional and other basic needs and in the calculation of corresponding cost, the experts or politicians who decide on a certain method must make value judgments (Hagendaars 1986).

Moreover, money income may not be a good measure of real consumption. During the Soviet period, most informal sector activities were illegal, and some residual stigma may induce households to under-report such income. Because the goal of poverty analysis is to consider people's real well-being, expenditure can be a better measure than income (Deaton and Muellbauer 1980). The definition of expenditure used in the Russian Longitudinal Monitoring Survey results reported below is more akin to consumption than to income, because it includes the imputed value of in-kind goods and services produced in the home, and received (for example, from employers).

Whether based on income or expenditures, monetary indicators of welfare can also cause problems for analysts trying to make comparisons across economic regimes. High inflation creates additional difficulties in measuring and comparing nominal aggregates. Another complication arises from the shift from a regime that relied on several different allocation mechanisms for goods and services, including queues and rationing with black markets. Searching and queuing imposed significant costs—although they were not evenly distributed. But price liberalization has largely eliminated scarcities and trade reform has increased the volume and quality of goods available. Consumer surveys reveal that shortages have gradually diminished over the transition, first in the metropolitan areas, and then in provincial and rural parts of the country. It has been argued that the net welfare gains now that queuing is no longer required have been significant, even where real incomes and consump-

tion fell (Roberts 1995). This observation, however, does not take into account distributional effects, which are obviously important in measuring poverty. Consumers on low incomes may actually prefer queuing at low prices because their opportunity costs of waiting are low. Even while money has become a much more meaningful indicator of welfare in Russia, comparisons over time are bound to be approximate. On balance the best approach is to base our assessments on household expenditure, mindful of the qualifications noted above. The next question is where to draw the poverty line against which to compare expenditures.

Drawing the Poverty Line

Drawing the official poverty line is in part a question of judgment and in part a matter of policy and politics. Thus it is important to bear in mind that the extent of measured poverty in the analyses here is sensitive to where the poverty line is drawn.[1]

Under Khruschev, a "minimum consumption budget" was developed (Sarkisyan and Kuznetsova 1967; see also Atkinson and Micklewright 1992), although official estimates of its ruble value as a poverty line were not published.[2] By January 1992, when extensive price liberalization was undertaken, the minimum consumption budget was manifestly unusable, because virtually the entire population had incomes below this standard (Mozhina 1992). A revised basket corresponding to World Health Organization guidelines was subsequently adopted, and the assumed high share spent on food (68.3 percent) was consistent with the structure of low-income household expenditures in the early years of transition, although future adjustment will be necessary to take account of changing relative prices—especially for services, which de Masi and Koen (1995) find extremely cheap relative to market economy standards. Table 1 compares the nominal value of the official poverty line to the average wages during the 1990s. The poverty line (per capita) has typically been only about half the average wage, suggesting that the threshold is quite austere. It is this poverty line, as calculated by the Ministry of Labor in Moscow, that we use in the analysis below.

Data Sources and Problems

Official censorship of poverty issues does not mean that information about incomes and expenditures did not exist in the USSR. The Family Budget Survey (FBS), instituted in 1922 (Dmitrichev 1992), is still the most widely cited source of information about poverty in Russia and other former Soviet republics. About 60,000 families made up the survey pool in 1970–85, which widened to 90,000 in 1988–90. In 1995 the Russian sample was approximately 50,000 families. At least through 1996, the survey was completed every quarter.

Although the FBS represents a rich potential source of data, the survey's methodology has been subject to extensive criticism, both internally and externally (Rimashevskaya

Table 1. *Official Poverty Lines and the Average Wage, 1980–96*
(rubles)

Year	*Minimum consumption basket*	*Official poverty line*	*Average wage*
1980	64.6	—	—
1990	93.3	—	—
1991	190.0	—	770
1992	—	4,282	16,071
1993		42,800	141,000
1994	—	145,400	354,000
1995	—	321,000	735,000
1996		379,000	944,000

— Not available.

Source: Estimates of the Soviet minimum consumption basket are for the whole year, based on press reports (see Braithwaite 1997). The official poverty line is the prozhiochniyy minimum, as calculated by the Ministry of Labor in Moscow. The yearly data are for December.

and others 1979; Shenfield 1983; Braithwaite and Heleniak 1989; and Atkinson and Micklewright 1992). A major concern is the sampling frame, which does not encompass the entire population but remains based on the so-called branch principle. Workers are drawn from enterprise rolls, with large and urban enterprises overrepresented and workers with seniority much more likely to be included given the material incentives to participate. The more workers in a family, the greater the likelihood that the family will be selected. A separate sample was set up for pensioners, but again, those retired from large enterprises are more likely to be included. The net result is that the lower part of the income distribution is not adequately represented. And the exclusion of certain occupational groups, such as senior bureaucrats, KGB officials, and military officers, raises some concerns about the upper end of the distribution (although this is a perennial problem in market economies as well). Nor does the design allow the calculation of sampling errors or confidence intervals, statistical techniques that allow an assessment of the reliability of results obtained from the survey.

Further, the interviewing and recording practices are questionable; the interviewer leaves a diary with the household and returns frequently to assist them in filling it out. This practice suggests the possibility of the interviewer simply creating and recording the data. The household is responsible for keeping both expenditure and income data. Before the transition there were severe penalties, including imprisonment, for illegally earned income from informal sources and from private capital or enterprise, and there was thus a very high degree of correspondence between stated household income and expenditures. For example, during the anti-alcohol campaign of the 1980s, households under-reported alcohol purchases to such an extent that the alcohol beverages weight in the consumer price index had to be imputed, using data from retail sales.

There are also serious problems associated with the analytical methodology currently used by the Goskomstat Rossii, the government statistical agency, particularly the calculations of monthly indicators of income and distribution.[3] As the figures reported in table 2 suggest, the official monthly estimates of poverty and distribution are not very plausible. Goskomstat's reported poverty head counts fluctuated markedly from month to month in 1993 and 1994. Nonetheless, the FBS was the basis for reports and analysis of poverty and distribution published by the Goskomstat.

Table 2. *Official Trends in Poverty and Distribution of Income in Russia, 1980–96*
(percent)

Year	*Head count*	*Poverty gap*	*Severity*	*Gini coefficient*
1980	11.3	2.34	0.70	27.60
1985	13.4	2.96	0.94	27.56
1989	11.0	2.24	0.65	26.49
1990	10.1	2.12	0.63	28.45
1991	11.4	2.03	0.56	26.54
1992				
January	30.2	7.25	2.45	23.93
June	23.1	6.59	2.68	28.70
Dec.	15.7	4.11	1.49	34.98
1993				
March	34.7	10.48	4.45	29.38
June	24.7	7.00	2.75	34.67
1994				
January	34.9			40.9 (annual)
May	16.4		—	—
Sept	20.5	—		—
1995				
January	33.0		—	38.1 (annual)
July	26.0		—	—
Dec	20.0		—	—
1996				
January	25.0	—	—	37.5 (annual)
July	21.0			—
Dec	18.0			—

— Not available.

Note: For 1980–91 the minimum consumption basket was used as the poverty line; for the period since, the subsistence minimum. The head count index shows the percentage of individuals in the population falling below the poverty line. The poverty gap index sums all the poverty gaps in the population, that is, the amount of money needed to bring all the poor up to the poverty line, as a share of GDP. The severity index gives greater weight to those individuals (households) furthest from the poverty line (see Ravallion 1992 for further details). The Gini coefficient is closer to zero when incomes (expenditures) are more equally distributed.

Source: Authors' estimates based on published Goskomstat sources.

In July 1992 Goskomstat Rossii implemented a new household survey, the Russian Longitudinal Monitoring Survey (RLMS), with technical and financial assistance from the World Bank and the U.S. Agency for International Development, that ran in parallel to FBS. The first round of the RLMS was nationally representative and involved 6,500 households. It was designed to meet scientific standards for a true probability sample (see Klugman and Braithwaite 1997 for details on the sampling method). The availability of this data has made it possible to undertake detailed analyses of poverty during the transition (World Bank 1995b; Klugman 1997; Glinskaya and Braithwaite forthcoming).

The demographics of the original RLMS sample compared favorably with the Soviet census that was carried out four years earlier (Foley 1997). The initial rounds achieved a high response rate (about 90 percent), although participation declined in later rounds. The original panel of households was replaced with a new panel after the fourth round. The household questionnaire collects data on expenditures, income, housing conditions, and ownership of land and assets. The individual questionnaire covers employment, use of time, migration, and anthropometry of young children.

Trends during the Transition

Since 1991, the number of poor households in Russia has risen sharply, and according to the RLMS, a record 35 percent of the population was living below the official poverty line by the end of 1995 (table 3). The share of individuals in poverty is higher (41 percent in 1995) than the share of households because poor households tend to be larger than average. Beyond the head-count measure, however, indicators suggest that the worst poverty occurred in 1993, when both the share of the popula

Table 3. *The Incidence, Depth, and Severity of Poverty among Households, 1992–95*
(percent)

Unit	*1992*	*1993*	*1994*	*1995*
Poor households[a]	25.2	31.9	26.8	35.0
Very poor households[b]	8.4	12.0	10.4	10.9
Depth	9.8	13.6	11.7	13.2
Severity	5.4	8.0	7.2	6.9
Head count for individuals	26.8	36.9	30.9	41.1

Note: Percentage of households (unless noted). Expenditure based.
a. Percent of households with expenditure below the poverty line.
b. Percent of households with expenditures less than half the poverty line.
Source: Rounds 1-6 of the RLMS; Foley (1997); Kolev (1996).

tion classified as very poor (households whose expenditures were less than half of the household-specific poverty line) and the severity of poverty peaked. By late 1995 about 11 percent of households in Russia were very poor, significantly higher than in 1992 although below the peak of 1993. Despite some abatement during 1994, the poverty head count at the end of 1995 was some 10 percentage points higher than it had been in mid-1992, and the depth and severity measures were also markedly higher.

These overall trends in poverty result from rapid declines in national income and changes in the distribution of income during the transition. Based on the early years (Milanovic 1995) estimated that an additional 1 percent decline in Russia's national income would lead to only 0.5 percent increment in the poverty head count. Decomposition of more recent trends in poverty suggests that the continued lack of significant real growth in 1995–97 accounts for most of the increase in poverty, and that household inequality has been fairly constant during the past two years. There are indications that a group of long-term poor may have emerged (Glinskaya and Braithwaite forthcoming).[4]

Who are the poor in contemporary Russia? About two-thirds of those living in poverty are in households where the head of the household is employed. They are primarily families with children, the unemployed, and single elderly people living alone. Nearly 85 percent of families with three or more children under 6 years of age were poor in 1994. Poverty, always associated with family size, has become increasingly concentrated in families with children as well as in households with an unemployed person. Thus while the overall head-count index increased by about 5 percentage points between 1994 and late 1995, in the same period poverty in families with one child rose by 14 points, and by more than 19 points in families with children under six years of age. Cross-tabulations show that in late 1995, consistent with the pattern that had emerged earlier in the transition, the incidence of poverty was almost 52 percent among children under 18, and an even higher 57.6 percent among children under six, compared with 41 percent for working age adults and 26.7 percent for the elderly. Controlling for other observed factors, the presence of one unemployed person in a household increased the poverty risk of the household by more than 15 percentage points above the national average.

Causes of Increased Poverty

Dynamic aspects of poverty were investigated between 1992 and 1994 through the RLMS panel. There were significant flows into and out of both poverty and severe poverty, even while the overall incidence of poverty was rising.

Table 4. *Changes in the Gini Coefficient for Money Income, Selected Countries, 1989 and 199*

Country	Gini coefficient 1989	Gini coefficient 1994
Bulgaria	25.0[a]	35.3
Estonia	27.2	38.6
Lithuania	27.5	37.0
Moldova	26.7	40.0
Poland	26.9	30.4
Romania	26.9	28.4
Russia	25.7	40.9

a. 1990.

Source: UNICEF TRANSMONEE, based on data reported by national statistical offices, UNICEF, Internatio Child Development Centre, Florence, Italy.

A key element of trends in poverty has been rising inequality. The disruption o the old system with its controlled wages and prices has clearly led to greater differe tiation in income-earning opportunities. Unlike the output shocks associated wi the transition, the disparity in income associated with changes in relative prices likely to persist. Table 4 suggests that the rise in Russia's Gini coefficient (a wide used measure of income inequality, with higher values indicating less equal distrib tion) is somewhat greater than that of most Eastern European countries, yet of similar order of magnitude to several former Soviet countries (including Eston Lithuania, and Moldova).

Measurements based on Goskomstat (1995) data suggest that the ratio betwe the highest and lowest income deciles widened dramatically, from 4.5 in 1991 t 15.1 in 1994, then moderated slightly to 13.5 in 1995. Goskomstat's calculate measures of inequality are significantly below other estimates, mainly for reaso related to sampling. RLMS calculations, not surprisingly, reveal higher levels of in equality both at the outset and through the transition. Given the weaknesses of th FBS in measuring income distribution, this finding seems more plausible than a ve sharp rise in inequality within the space of a few months. The RLMS-based Gi coefficient stood at 49.2 by mid-1993, then declined slightly to 48.4 by the end o 1995. Trends in the decile distribution based on the RLMS reveal that the top decile share of total expenditure rose from about 33 to more than 37 percent between 199 and 1995. The relative share of each of the bottom five deciles also increased, albe slightly from very low levels, over the same period. Hence in relative terms the con sumption of the former middle class appears to have been squeezed most. This i consistent with Milanovic's (1995) observation that the income of the former middl class (which includes clerical staff, production workers, teachers, administrators, an doctors) had declined significantly. The changes in relative wages analyzed abov help to explain these shifts.

The Underlying Mechanisms

What mechanisms underlie these developments? In this section we attempt to link the risk factors associated with poverty in contemporary Russia to changes in the labor market and in the extent and direction of public transfers.

Adjustments in the Labor Market

According to Goskomstat data, wages, salaries, and entrepreneurial income amounted to almost 80 percent of total personal income in 1995 (public and private transfers accounted for the remaining 20 percent). Changes in the level as well as in the relative share and distribution of these components significantly affect household welfare. Labor market adjustments have been extensive in the wake of the price and wage liberalization involved in the move to a market system (Commander and Yemtsov 1997). Aggregate declines in the demand for labor as a result of the drop in output have led to rising unemployment, albeit at a slower rate and to a lesser extent than originally expected by external observers, including the World Bank. This is partly because significant cuts in real wages have accompanied the decline in demand. Average real earnings during 1992–95 were about half the level of the 1980s (table 5).

Various factors appear to determine labor market rewards. Education and skill levels appear to be factors in determining whether someone is poor. The results of the econometric analysis of the RLMS data indicate that higher educational attainments are associated with a somewhat lower risk of poverty: for example, the risk is reduced by 9 percent for households headed by university graduates. Households

Table 5. *Trends in Real Benefits and Wages, 1989–96*

Year	*Average pension*	*Minimum pension*	*Family allowance*	*Minimum wage*	*Average wage*
1989	82	68		109	70
1991	73	85	—	85	67
1992	38	43		43	36
1993	43	37		37	50
1994	55	49	27	27	49
1995	49	32	12	12	35
1996	48	33	20	18	37

— Not available.

Note: Fourth quarter 1991 equals 100. For the years 1992–96 inclusive, the reference month is January; for 1991 first quarter; and for 1989 annual.

a. It is not possible to give a single figure for the period before 1994 because of the variety of benefits (see text). The base was calculated using an estimate based on the minimum wage (to which a number of benefits were linked).

Source: Braithwaite (1997), table 2.1; Ministry of Labor, Moscow.

that are headed by the unskilled, by clerks, and agricultural workers are at higher risk of poverty. Finally, and perhaps less expected, the number of individuals in a household engaged in primary employment (that is, a main job) does not significantly affect the risk of poverty. Participation in secondary employment (often in the informal sector) does reduce the expected probability of poverty, however, by about 11 percent.

The liberalization of wage decisions was expected to affect the distribution of income significantly. Before the transition the share of wages in total personal income was almost 75 percent, which is high by international standards. For example, in 1989 wages accounted for only 59 percent of total personal income in the United States (U.S. Department of Commerce 1991). Increasing inequality in the distribution of total income in Russia has been largely driven by the decline in the share of wages in total personal income and the corresponding rise in the contribution of entrepreneurial and capital income (figure 1). The drop in the wage share is quite plausible in the context of the previous regime's ban on virtually all types of entrepreneurial earnings and private capital investments. In 1990 wages and entrepreneurial income (including capital) contributed 74.1 and 12.9 percent respectively to total monetary income. By 1995 the combined share of entrepreneurial (38.6) and capital (5.2 percent) income outstripped the 39.5 percent share contributed by wage

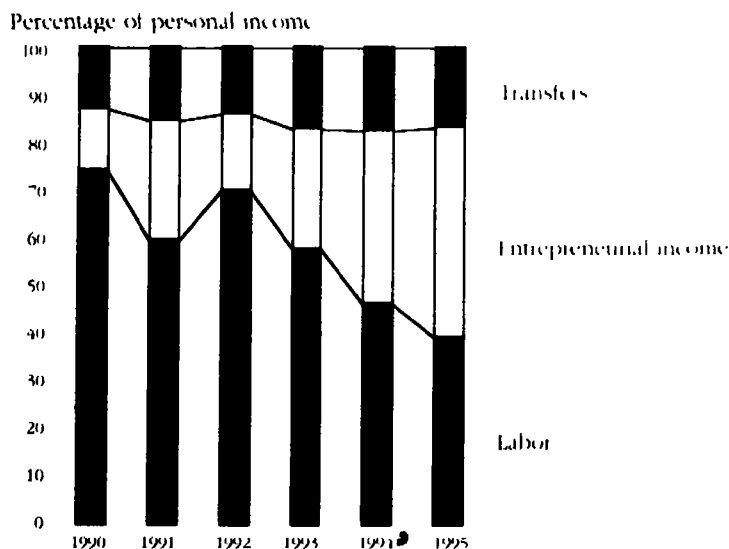
Figure 1. *Structure of Personal Income, 1990–95*

Note: Entrepreneurial income includes capital income, which amounted to 5.2 percent in 1995.
Source: Goskomstat (1996).

Table 6. *Selected Wage Indices Relative to National Average Wage, 1990–95*

Sector	*1990*	*1991*	*1992*	*1993*	*1994*	*1995*
Industry	103	111	118	108	104	111
Agriculture	95	84	66	61	50	43
Construction	124	127	134	133	129	131
Transport	115	120	146	151	150	152
Commerce	85	91	91	107	123	117
Science/R&D	113	90	64	68	78	78
Finance	135	180	204	243	208	. .

Note: National average is 100.
Source: Goskomstat (1996).

income. These changes reflect the dramatic erosion in the average real wage as productivity of employed labor has fallen over the period. As in other transition countries, such as Bulgaria (Milanovic 1995), the growing share of entrepreneurial and capital income has driven the overall shift to a less equal distribution of income.

Higher overall wage inequality has been accompanied by sectoral shifts in the economy and changing sectoral wage relativities. Table 6 suggests that workers in the industrial, construction, transport, commerce, and especially, financial, sectors have done relatively better than average. The most marked deterioration occurred in agriculture, in science, and in research and development. The decline in agricultural wages reflects the withdrawal of state subsidies combined with declines in output and only limited adjustment in employment. Additionally, Russia's delay in undertaking comprehensive land privatization likely hurt agricultural laborers. In Armenia, by contrast, swift privatization was associated with rates of rural poverty that were significantly less severe than in urban areas (Braithwaite 1995). The official wage figures support the findings of the RLMS that the incidence of poverty among households headed by skilled agricultural workers in late 1995 was the highest among all reported job occupations, at 64.[illegible] percent, even higher than unskilled workers ([illegible] percent) and far higher than the national rate of about 35 percent.

In the face of inherited overstaffing and falling demand for labor, employers have laid off increasing numbers of workers or reduced their hours of work. By mid-1996 registered unemployment stood at about [illegible] percent of the labor force, although International Labour Organization methodology, which includes all unemployed individuals whether or not they are formally registered with the Federal Employment Service, puts it at about 9 percent. Blue-collar workers have borne the brunt of the decline in labor demand. Hours were cut for about 5 million workers in 1994, and another 7.4 million were placed on involuntary leave. During 1993 and 1994 only 40 percent of the work force was paid in full and on time. If wages are paid late and consequently are eroded by inflation or are not received at all, earnings figures can be quite misleading. Multivariate analysis of the RLMS suggests that households report-

ing wage arrears are 11 percent more likely to be poor. Overall, however, firms a pear to prefer to retain workers, a preference that can be traced to insider influence the firm level, as well as to the limited social safety net for displaced workers.

The incidence of poverty among the unemployed is high; in late-1995 about h of households with an unemployed member were poor. An additional unemploy person in the household raised the probability of a family being in poverty by abo 9 percentage points. The Federal Employment Service, which is responsible for i come support and labor market programs for the unemployed, reaches only a se ment of its target group. Commander and Yemtsov (1997) found that the majori of the unemployed do in fact have some marginal employment and that, as expecte there was a strong positive link between marginal employment and educatio quitting a previous job (as opposed to being laid off), and residence in Moscow St. Petersburg.

Increasing regional disparities—apparently caused by changing relative pric along with reductions in the subsidies to the producer—are a critical dimensi of recent labor market trends. Industrial areas in central Russia and the Nor Caucasus have been especially hard-hit, while regions with abundant natu resources, such as Yakutia, and commercial centers such as Moscow and St. Peter burg have suffered less. Trade liberalization has had differential impacts becau of the sectoral concentration of economic activity. The changing structure demand has also had an uneven effect across regions, for the same reason. T most obvious example is the impact on areas that were previously dependent o military production.

Indeed, there is an inverse relation between industrial output and register unemployment, with oblasts that did not suffer from high unemployment wage arrears tending to have experienced either substantial real wage cuts relatively less industrial decline. Theory and the empirical evidence suggest th regional disparities in unemployment are likely to persist, despite signs of wag flexibility and an emerging inverse association at the regional level betwe changes to wages and unemployment. Commander and Yemtsov (1997) ha identified a geographic mismatch in the distribution of the unemployed and th availability of jobs that is associated with severe constraints to labor mobili (particularly lack of housing).

The average duration of unemployment is increasing but is still relatively short the end of 1994 labor force surveys suggested that about 60 percent of the unem ployed found a primary job within six months, while 23 percent were out of wor for longer than a year. These relatively short episodes indicate that gross flows i Russia's labor market were large, especially compared with most European transitio economies where unemployment was generally higher, while hiring is increasingl concentrated in the more flexible nonstate and informal sectors (Commander an Yemtsov 1997).

The government's social policy response appears to have failed to stem the rise in poverty. The formal system of social protection has become increasingly inadequate for many vulnerable population groups, with the important exception of those who receive old-age pensions. We have examined the impact of public transfers by analyzing the incidence of various benefits at different dates using the RLMS as well as movements in the real official level of benefits.

During the Soviet period, the system of social support was based upon low administered prices for food, rent, household utilities, and other basic goods and services, along with a guaranteed job. This inherited system of support for households was not easily adapted to the introduction of market principles. Nor was the structure of social protection adequate to deal with the type and scale of needs created by the transition, even though this protection included both social insurance (pensions and unemployment benefits) and social assistance programs (including family allowances). Sources of financing have been squeezed because social insurance is largely financed through payroll taxes, which have fallen, whereas social assistance is largely the responsibility of local authorities whose economic fortunes have varied enormously.[5]

The government allocated about 4.5 percent of total budget expenditures for social transfers in 1995, most of which was provided by regional budgets. The real value of transfers has been seriously eroded, however, as GDP has declined (see table 5). The average pension, which has hovered at about half the level of late 1991, has nevertheless been relatively better protected than both the average wage and other benefits. The minimum pension was also badly affected during the same period, falling to nearly one-third of its level in 1991.

During the Soviet period, family allowances were paid to single mothers and to large families. A new system of universal child allowances evolved in the wake of the price liberalization; a complicated system with some 18 different benefits for families with children. Because most of these were set as some proportion of the minimum wage, however, lagged indexation significantly eroded their value (World Bank 1994). In 1994 the system was simplified with the introduction of a unified benefit. But the real value remained very low, at only about one-fifth of the pretransition levels.

Increasing regional disparities in poverty rates can be attributed in part to the effects of restructuring in regions with different economic bases, but also to the devolution of responsibility for financing social assistance to local authorities. Disparities in resources and incomes have constrained local authorities, who also have responsibility for financing and delivering the bulk of education and health services as well as subsidizing housing and domestic utilities. Thus decentralization has contributed to the trend toward regional inequality. Analysis of federal fiscal transfers shows that the impact is not progressive (World Bank 1995a). The introduction of a new fed-

eral transfer mechanism in 1994 was indeed perversely disequalizing, apparently because the transfer was based on the oblast's own assessment of need derived from the current level of service, which tended to be higher in wealthier areas. Evidence suggests that health and education expenditures have been relatively protected against budget cuts but that social assistance has been cut disproportionately. Stewart (1996) found that no correlation between a locality's official poverty head count in 1994 and the amount it spent on social assistance, meaning that areas with greater relative need spend less on social assistance.

Because pension payments are closely related to previous employment and levels of income, these transfers are regressive (World Bank 1995b). Pensions appear to be an effective means of intergenerational transfer in that reported poverty among the elderly has been kept consistently below the national average since 1992. This pattern mirrors that of an increasing number of western industrialized countries in which public pension systems have signifcantly reduced, and in some countries virtually eliminated, old-age poverty (Scherer 1997). Among pensioners who are not poor (table 7), pensions represent more than half their income. For the elderly living in very poor households, pensions contribute an even larger share of their income

Table 7. *Coverage and Significance of Public Transfers, 1994 and 1995*
(percent)

	Very poor		*Poor*		*Non poor*	
Type of transfer	*Receiving the benefit*	*Share of recipient household income*	*Receiving the benefit*	*Share of recipient household income*	*Receiving the benefit*	*Share of recipient household income*
Family allowance						
1994	28.8	23.6	32.4	14.5	[illegible]	[illegible]
1995	29.3	16.3	32.3	13.1	[illegible]	[illegible]
Pension						
1994	40.3	75.0	41.0	66.9	48	[illegible]
1995	36.8	61.5	39.7	54	54.0	[illegible]
Unemployment benefit						
1994	0.8	21.7	0.4	17.8	0.3	[illegible]
1995	2.1	37.6	2.6	19.6	0.9	[illegible]
Local social assistance[a]	10.4	9.6	10.4	9.6	14.5	[illegible]
Housing subsidy[b]	3.3	7.5	3.6	7.5	6.4	[illegible]
Scholarship	5.2	17.8	6.2	18.2	6.7	[illegible]
All transfers						
1994	66.8	58.5	70.9	48.4	71.4	[illegible]
1995	61.6	43.3	67.7	40.3	70.0	[illegible]

a. Data all for 1994, first quarter.
b. Information is for December 1995 and January 1996.
Source: Foley and Klugman (1997); Kolev (1996)

As suggested above, the Federal Employment Service has been largely ineffective in providing social protection for the unemployed. Indeed, table 7 reveals that in late 1995, when about 13 percent of poor households had an unemployed member, less than 3 percent of the poor received unemployment benefits. Moreover, given its minimal level (see table 5), the contribution to income is only significant for the very poor.

According to the RIMS, from 1992 to late 1995, public transfers accounted for almost 30 percent of total household income; the average contribution to the income of *recipient* households was higher—an estimated 42 percent in late 1995. Important trends during this period can be observed regarding the relative importance of different transfers (see table 7). Pensions remain the most widely received benefit. Unemployment benefits, although still very limited in their coverage, are going to an increasing number of the poor and accounting for a growing share of their income. Coverage of family allowances was steady for the poor, but dropped for the nonpoor, probably because families that are not needy may not bother to apply. Such benefits also declined as a share of income for poor and very poor households, given the erosion of their real value and the delays in payment.

Although the coverage of public transfers across the population is fairly extensive, there are some perhaps unintended results. For example, most of the poor and very poor are eligible for transfers, and yet they do not receive benefits (table 8). Nearly four out of ten very poor households and one out of three poor households did not receive any transfers. Yet the majority of nonpoor families *did* receive public transfers. Most poor households not receiving benefits are in fact eligible for them. These errors of exclusion appear to be administrative problems, fiscal ones (that is, local authorities do not have sufficient funds to pay benefits), or both. Inadequate information about eligibility and lack of access to local offices may contribute to low take-up rates. The number of those receiving benefits to which they are not entitled (er-

Table 8 *Errors in Targeting of Public Transfers, 1995*

Category	Very poor	Poor	Non poor
Did not receive benefits	39.8	33.0	30.9
Did not receive benefits although eligible	36.4	29.[illegible]	22.8
Not eligible to receive benefits	6.2	6.1	13.5
Received benefits although ineligible	0.8	1.5	3.9

Note: Public transfers considered here are limited to pensions, unemployment benefits, scholarships, and family allowances, based on the RIMS.

Source: Kolev (1996)

rors of inclusion) appears to be relatively limited. These overall patterns have persisted throughout the transition.

Only about one in five Russian households rely solely on income from employment and transfers from the formal system of social protection. Private interhousehold transfers have been investigated by Cox, Eser, and Jimenez (1997), who found that private transfers in Russia are large and responsive to the socioeconomic characteristics of the household. Private networks are extensive, with four out of ten Russian households acting as donors, recipients, or both. Longitudinal analysis shows that this behavior has persisted throughout the transition, in contrast to Poland, for example, where private transfers diminished as economic conditions worsened. For the RLMS sample as a whole, private transfers average about 5 percent of household income; for net recipients the contribution was much higher, amounting to about 20 percent of household income.

Overall, private transfers appear to flow to such vulnerable groups as younger families, female-headed households, and households affected by unemployment. Households participating in private transfers tend to be better off than those who do not participate, although whether such transfers have an equalizing effect on the distribution of income is unclear. Although private transfers tend to be directed from better-off to poorer groups, the probability of receiving a transfer declines only slightly as earned income increases. Empirical investigation reveals that the theoretical concern about public transfers crowding out private transfers are not warranted: private and public transfers appear to be complementary.

In light of this evidence, it is perhaps not surprising that overall attitudes toward government-provided social protection are mixed. On the one hand, surveys suggest an increasing degree of self-reliance. The vast majority say they would first rely upon themselves, and then friends and family, for help in time of need. Yet most people still expect the government to make good on the perceived right to employment for every able-bodied person. Moreover, the expressed needs for social support from the government, as measured in nationally representative public opinion surveys, are significant (Zubova and Kovaleva 1997).

Conclusion

Analysis of household survey data shows that Russia's transition from a command economy has been associated with a significant worsening of household and individual well-being. Although some indicators suggest that the worst may have passed in 1993, the numbers of individuals and households below the poverty line rose again in 1995, despite signs of economic recovery.

Russia is not unique among economies in transition in experiencing worse and more prolonged poverty than many expected at the outset. As elsewhere, aggregate

trends in poverty result from changes in output and income distribution. In both respects, however, the Russian experience has been relatively adverse. The lack of general consensus on an appropriate economic and social reform strategy delayed sustained stabilization and prolonged the period of high inflation and recession. The associated collapse of public revenue has severely strained the government's capacity to finance social transfers, as well as essential public goods and services

The publicly financed system of pensions has kept the rate of poverty among the elderly consistently below the national average. Yet for those outside the formal payroll-based system of social insurance, in particular such families with children, social assistance is ad hoc and limited. The same is true for the increasing number of households affected by unemployment, suggesting that the ongoing European debate about the social exclusion of marginalized groups may be relevant to Russia in the longer term. Several factors—low skills, type of employment, regional location, family structure, and benefit rules—appear to be interacting to marginalize some groups in society. At the same time, as elsewhere in transition countries and the West, popular demands for more comprehensive social assistance raise concerns about welfare dependency.

The emerging economic recovery does not yet appear to have begun to turn around poverty, according to the aggregate poverty measures. In this sense the experience of Russia might be analogous to that of Poland, where several years of positive growth occurred before poverty began to decline (World Bank 1995b). The extent of poverty in Russia — in 1995 some 11 percent of the population was very poor—suggests that the resumption of growth could take longer to reduce poverty. Moreover, restructuring is needed in several sectors of the economy, such as coal, metallurgy, chemicals, machine-building, and agriculture. Business reorganizations will likely entail a significant number of layoffs, thereby increasing unemployment and possibly poverty as well

The interest of the international community in the social impact of Russia's transition lies, at least in part, in its impact on the political sustainability of continued economic and democratic reform. Public opinion surveys are a potentially rich source of information in this regard (Zubova and Kovaleva 1997). In Russia, as elsewhere, subjective poverty standards are much higher than the officially established absolute poverty lines, so a much higher share of the population regards itself as poor. Belief in an egalitarian distribution of income remains persistent and widespread. People tend to attribute poverty to economic and structural causes, such as unemployment and lack of educational opportunities, rather than to individual characteristics. Russian society in general is rather pessimistic about the future, and the assessments of the poor are especially negative. These results suggest that even if poverty rates improve in the short term, widespread dissatisfaction with the process of economic and social policy reform may still persist for some time.

Notes

Jeni Klugman is with the Research School of Social Sciences at the Australian National University and is an associate fellow to the International Child Development Centre of UNICEF, Florence. She is presently on leave from the Europe and Central Asia Regional Office of the World Bank. Jeanine Braithwaite is an economist in the Poverty Reduction and Economic Management Network of the World Bank.

1. The poverty headcounts are not overly sensitive to shifts in the threshold, however. If, in late 1995, the poverty line had been moved up by 10 percent, the poverty incidence rises by about [illegible] percent; a 10 percent lower threshold reduced the poverty headcount estimate by 5 percent (Kolev 1996).

2. The budget was intended to be an absolute measure of minimally acceptable consumption under a socialist system, rather than an absolute measure of poverty per se (since, after all, poverty did not officially exist). Thus the Soviet standard for protein consumption was twice the level recommended in the United States (Lane, Marston, and Welsh 1987).

3. The published monthly income distribution data are based on the FBS, but are not actually summary totals from that survey. The FBS survey forms are collected and processed quarterly. The monthly income distribution data are synthetically generated by Goskomstat, using a previously tabulated FBS distribution as a historical template for the variance and grossing up the distribution by presumed increases in the mean. Average income is assumed to grow at the rate of such various monthly macroeconomic indicators as average wages or the wage funds of large state-owned enterprises. Generally, such an estimating methodology would be expected to lead to understatements in measures of income dispersion.

4. A decomposition breaks down changes in the poverty headcount among three components: growth, redistribution, and a residual component that is ill-defined (Ravallion and Datt 1991).

5. The Employment Fund, which is responsible for financing income support for the unemployed as well as programs for the employed, faces both problems. The financing base is a 2 percent payroll tax, the bulk of which is retained at the oblast level, limiting scope for any redistribution of funds to finance unemployment benefits in badly hit regions.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Atkinson, Anthony B., and John Micklewright. 1992. *Economic Transformation in Eastern Europe and the Distribution of Income*. Cambridge, U.K.: Cambridge University Press.

Braithwaite, Jeanine. 1995. "Armenia: A Poverty Profile." PSP Discussion Paper 80. World Bank, Poverty and Social Policy Department, Washington, D.C. Processed.

———. 1997. "The Old and New Poor in Russia." In Klugman (1997).

Braithwaite, Jeanine, and Timothy E. Heleniak. 1989. "Social Welfare in the USSR. The Income Recipient Distribution." Center for International Research, U.S. Bureau of the Census, Washington, D.C. Processed.

Commander, Simon, and R. Yemtsov. 1997. "Russian Unemployment. Its Magnitude, Characteristics, and Regional Dimensions." In Klugman (1997).

Cox, Daniel, Zekeriya Eser, and Emmanuel Jimenez. 1997. "Family Safety Nets during Economic Transition: A Study of Inter-Household Transfers in Russia." In Klugman (1997).

Popov, Vladimir. 1996. "A Russian Puzzle: What Makes the Russian Transformation a Specia Case?" Research for Action 29. United Nations University/World Institute for Developmen Economics Research, Helsinki. Processed.

Ravallion, Martin. 1992. *Poverty Comparisons: A Guide to Concepts and Methods.* LSMS Workin Paper 88. Washington D.C.: World Bank.

Ravallion, Martin, and Guarav Datt. 1991. *Growth and Redistribution Components of Changes in Poverty Measures: A Decomposition with Applications to Brazil and India in the 1980s* LSMS Working Paper 83. Washington D.C.: World Bank.

Rimashevskaya, N. M., and others. 1979. *Potrebnosti, dokhody, potreblenıye* ["Needs, Income, and Consumption"; not available in translation]. Moscow: Nauka.

Roberts, Bryan. 1995. "Price Liberalization, Market Power, and Social Welfare in Transition Economies." University of Miami, Miami, Florida. Processed.

Sarkisyan, G. P., and N. P. Kuznetsova. 1967. *Potrebnosti i dokhod sem'i: uroven', struktura, perspektivy* ["Needs and Incomes of Families: Level, Structure, and Perspectives"; not available in translation.] Moscow: Nauka.

Scherer, Peter. 1997. "The New Social Policy Agenda." Organization for Economic Cooperation and Development, Social Policy Division, Paris. Processed.

Shenfield, Stephen. 1983. "A Note on Data Quality in the Soviet Family Budget Survey." *Soviet Studies* 35(4, October):561–68.

Stewart, Kitty. 1996. "Regional Disparities in Social Expenditure in Russia." World Bank, Europe and Central Asia Regional Office, Washington, D.C. Processed.

U.S. Department of Commerce, Economics and Statistics Administration. 1991. *Statistical Abstract of the United States,* 11th ed. Washington, D.C.: Government Printing Office.

World Bank. 1994. "Social Protection during the Transition and Beyond" vol. 1. Report 11 48 RU. Europe and Central Asia Department, Washington, D.C. Processed.

———. 1995a. "Fiscal Management in the Russian Federation." Europe and Central Asia Department, Washington, D.C. Processed.

———. 1995b. "Poverty in Russia: An Assessment." Report 14110-RU. Europe and Central Asia Department, Washington, D. C. Processed.

Zubova, Larissa, and Natalia Kovaleva. 1997. "Public Opinion about Social Issues." In Klugman (1997).

How Bad Is Unemployment in Tunisia? Assessing Labor Market Efficiency in a Developing Country

Martin Rama

unisia's unemployment rate has been among the highest in the world for almost two lecades. This article claims that such a high rate reflects measurement problems rather han labor market inefficiency. After discussing the reasons why unemployment rates may ot be comparable across countries and reviewing the tools that are available to analyze nemployment in a specific country, the article provides four pieces of evidence to substantiate its claim. Two of them relate to the criteria used in Tunisia to measure unemployment and the way these criteria have changed over time. Two use records on the number f active job seekers and vacancies as reported to the official employment agency. Together, his body of evidence suggests that unemployment has declined steadily over time and emains an issue for first-time job seekers only.

unisia's unemployment rate is unusually high. Drawing from the 1989 labor force urvey, official estimates put that rate at 15.3 percent. Unofficial estimates for more ecent years, based on labor force and employment forecasts, yield rates as high as 16 r 17 percent. Except for economies in transition to a market orientation, only a andful of countries—Algeria, Barbados, Finland, Jordan, the Seychelles, South Afica, Spain, and Trinidad and Tobago—report similar or higher rates (ILO, 1996). Whether unemployment is as high in Tunisia as the estimates suggest is an issue that he population census of 1994 could clarify, but it may take some time before the ensus data are processed, and it is not obvious that the computations will rely on an nternationally comparable definition of unemployment. In the meantime, the country onfronts major policy reforms, including the implementation of a free trade agreement with the European Union and the privatization and downsizing of state-owned nterprises. The problem is that the consequences of these reforms depend on how fficient the labor market actually is.

World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 59–77

If the available unemployment figures are correct, the labor market is very inefficient, which may well force the government to reconsider its policy choices. On the one hand, if it seems likely that those workers who lose their jobs as a consequence of trade liberalization and ownership divestiture will be unable to find new jobs in competitive sectors of the economy, or that they might manage to do so only after a long period of unemployment, the government presumably will hesitate to implement the reforms. On the other hand, the high unemployment rate could indicate significant distortions in the labor market that require a radical reform of labor market policies and institutions. In general, the almost automatic reaction to a high and stable unemployment rate is to aim at minimum wage and collective bargaining mechanisms that could be pricing workers out of jobs.

Assessing the efficiency of the labor market is seldom an easy task, even in industrial countries. Conceptually, the issue seems quite simple: the labor market is considered inefficient when a large number of individuals who would be willing to work (or to work more) at the prevailing wage rate are unable to do so. In practice, however, both work and willingness to work can be measured in different ways. For instance, it has been argued that Japan's unemployment rate, which is the lowest among industrial countries, is biased downward because some individuals who are not working are counted as employed, while they would be considered unemployed elsewhere (Hashimoto 1993). Similarly, Spain's unemployment rate, the highest in the industrial world, is said to be overestimated because many employed workers are not registered with the social security system (for conflicting views on the magnitude of this bias, see Franks 1994 and Blanchard and Jimeno 1995).

Measurement problems are magnified in the case of developing countries because self-employment and work within the family are much more prevalent than in industrial countries. As a result, the concept of willingness to work "at the prevailing wage rate" loses some of its meaning. For instance, for many women the distinction between working on a family farm (therefore, being a "worker") and taking care of the home (hence being "inactive") is tenuous, as is the distinction between being unemployed and working in the informal sector for many men. Similarly, it is unclear whether a young high school graduate looking for a job in the formal sector and meanwhile refusing to work in the informal sector should be considered unemployed or rather inactive. When all of these problems are taken into account, it becomes obvious the unemployment rate can be measured in more than one way (World Bank 1995c, chapter 3). It is therefore important to use the same measure when making comparisons across countries or over time for the same country.

The credibility of any assessment of labor market efficiency will vary substantially depending on the quality of the data available for the analysis. For example, microeconomic data on living standards may produce a detailed profile of the unemployed that could help explain why they are out of work. Data on earnings and individual characteristics of workers (such as education, experience, and the like)

and their jobs can be used to evaluate whether different sectors pay the same wage to similar workers, which in turn could help identify the sources of labor market inefficiency. But in many cases these data are not available, and governments cannot postpone key policy choices until they are collected and processed. The challenge is therefore to assess labor market efficiency based on the kind of data usually published by the statistical offices of developing countries. In this respect, focusing on a country like Tunisia represents an even bigger challenge. Nowhere in the world is the paucity of labor market statistics more acute than in the Middle East and North Africa region (Rama 1997). If a credible assessment of labor market efficiency is feasible in Tunisia, then it should be feasible in other developing countries as well.

This article argues that Tunisia's labor market is not nearly as inefficient as the unemployment figures suggest. Four pieces of evidence substantiate this claim. Each of them is partial and subject to criticism, but they all point in the same direction. The first concerns the criteria used to measure the labor force, and the way these criteria have changed over time. The second relates to the methods used to forecast the size of the labor force and total employment since the last labor force survey. The third relies on the number of active job seekers and the number of available vacancies, as reported to the official employment agency. The last uses these records to compare the efficiency of the matching process between job seekers and vacancies in Tunisia with that of other countries.

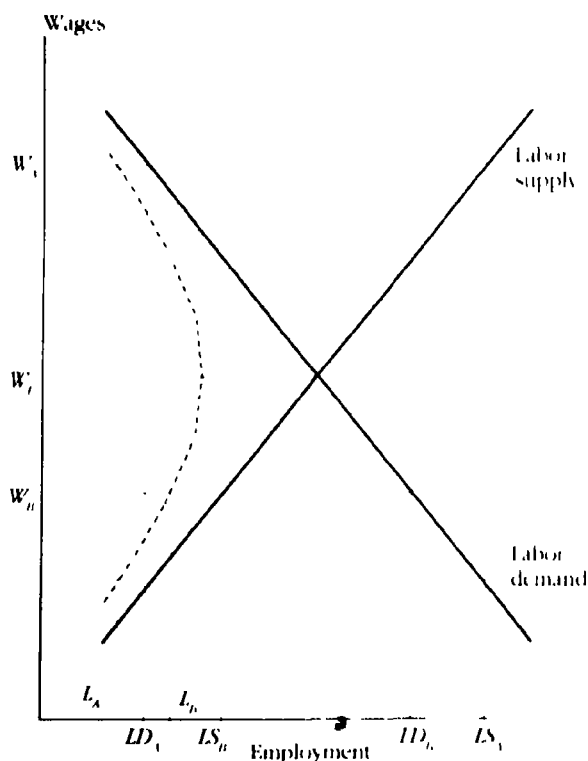
The same conclusion cannot be generalized to other developing countries with reportedly high unemployment rates, however. The article highlights the problems surrounding unemployment figures in developing countries, and suggests possible ways to understand what these figures really mean. But it does not claim that the biases observed in Tunisia plague the official figures of other developing countries as well. It only shows that biases of this sort may exist and may lead to inappropriate policy recommendations. Because the nature of any measurement biases is likely to vary from country to country, it is important to make the most of the labor market data available in each country. The four pieces of evidence examined here illustrate the broader proposition.

A Conceptual Framework

One of the problems in deciding whether an unemployment rate is actually "high" or "low" occurs because the most popular model of the labor market sets the equilibrium unemployment rate at zero. In practice, however, even the tightest labor market displays a strictly positive unemployment rate. In the basic model, shown in figure 1, as wages rise (on the vertical axis), more people are willing to seek jobs. This is shown by the increase in the labor supply (on the horizontal axis) along the upward sloping line. Conversely, when wages are extremely low, employers are willing

to hire large numbers of workers. But as wages rise, labor demand declines along the downward-sloping line. At the wage level W_F the number of jobs offered equals the number of workers seeking employment. Labor market equilibrium is thus characterized by no unemployment at all.

Within this simple model, the only reason why unemployment could be "high" (strictly speaking, positive) is because wages are misaligned. Assume that a minimum wage or a union-negotiated increase pushes the wage level up (in figure 1, from W_I to W_A). In this case, LS_A individuals would be willing to work, but only LD_A would find jobs, because wages would be so high that employers could not afford to hire many workers. This distorted equilibrium is what comes to the minds of many economists when data reveal a high unemployment rate. But in developing countries, usually characterized by a large informal sector, this interpretation is problematic. Indeed, most of the potential sources of wage misalignment are irrelevant in the informal sector: minimum wages are not enforced and unions are almost nonexistent. The

Figure 1. *A Basic Model of the Labor Market*

Source: Author's calculations.

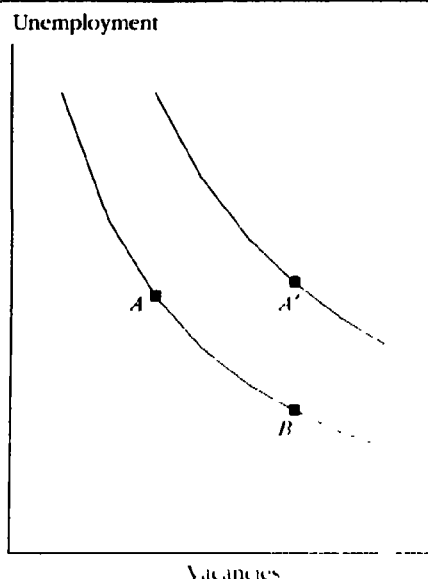
labor demand curve in figure 1 corresponds to the formal sector only. As wages in this sector rise, some workers lose their jobs. But instead of being forced into unemployment, they shift to the informal sector.

There are several ways of introducing equilibrium unemployment in simple models of the labor market like the one represented in figure 1. One assumes that workers in the informal sector cannot join the formal sector without previously spending some time unemployed. In the context of development economics, this assumption has been justified on the grounds that the formal sector is mostly urban. Agricultural workers choose between staying in rural areas and migrating to urban centers, where they may either find a job in the modern sector or remain unemployed. The equilibrium unemployment rate is the one that equalizes the expected payoffs of these two alternatives (Harris and Todaro 1970). That the informal sector is quite large in urban areas makes this explanation of unemployment unconvincing, however. Rural migrants could indeed work in the urban informal sector while waiting for openings in the formal sector.

A second way of introducing unemployment consists of assuming frictions in the matching process between job seekers and vacancies. The labor market is actually characterized by large flows in and out of employment. There are always individuals who quit their jobs, are fired, retire, or die, while newcomers are constantly joining the labor force. It takes time for job seekers to learn about openings that are suitable to them and for firms to find candidates who meet their needs. So even if the supply of labor were equal to demand at the aggregate level, there would always be some individuals in search of a job and some firms in search of the appropriate workers (Pissarides 1985; Blanchard and Diamond 1989).

Because of the frictions in the matching process, actual employment is less than both labor supply and labor demand, as represented by the broken curve in figure 1. For instance, when the wage level is high (equal to W_A), employment is given by L_A and there are simultaneously $LD_A - L_A$ unfilled vacancies and $LS_A - L_A$ unsuccessful job seekers. Conversely, when the wage level is low (equal to W_B), the employment level is L_B and there are simultaneously $LD_B - L_B$ unfilled vacancies and $LS_B - L_B$ unsuccessful job seekers. Although these two situations may look alike in terms of employment, they are radically different in terms of their ratios of vacancies to unemployment. Job seekers have better prospects of finding a job in the second situation, so the time during which they are unemployed should be shorter. More generally, the joint behavior of unemployment and vacancies can provide more information on the functioning of the labor market than the mere unemployment rate does.

A convenient representation of the joint behavior of unemployment and vacancies is provided by the Beveridge curve (figure 2). The curve on the left is based on the diagram in figure 1. Point A in this curve corresponds to the wage level W_A, while point B corresponds to the wage level W_B. The other curve displays a higher number of unsuccessful job seekers for any number of vacancies, or equivalently, a higher number of vacancies for any number of unemployed persons. These two curves are

Figure 2. *The Beveridge Curve*

Source: Author's calculations.

useful to illustrate the difference between movements *along* the Beveridge curve and *shifts* of the Beveridge curve. A movement from point A to B indicates that the labor market has tightened, without changing its overall efficiency. Business cycles are characterized by back-and-forth movements of this kind. A movement from point A to point A′, in turn, reflects a more inefficient labor market. That is, the same number of job seekers remain unemployed despite a larger number of unfilled vacancies. For instance, it has been argued that labor market policies and institutions in Europe are such that the temporary increase in unemployment created by the oil shocks (from B to A) has reduced labor market efficiency and become a more permanent situation (from A to A′, rather than back to B).

Another convenient tool to assess the importance of labor market frictions is provided by matching functions. These functions link the number of new placements, H, to the number of job seekers, U, and the number of vacancies, V. In a well functioning labor market, H should increase when more individuals are actively searching for a job and when more vacant posts are available as well. In analytical terms the number of placements is represented as follows:

$$H = h(U, V)$$

where $h(.)$ is the matching function. The shape of this function has been investi-

gated for several industrial countries. The econometric results show that, in general, if the number of job seekers and the number of offers doubled, placements would roughly double too. In technical terms, the *h(.)* function is characterized by constant or near-to-constant returns to scale. The results also show that both arguments of the function matter. Even in cases of massive job destruction, as in the transition economies, the stock of the unemployed remains an important determinant of the flow of hirings (see Burda 1993, on the former East Germany). This is in sharp contrast with the frictionless labor market shown in figure 1. Indeed, in that model, when the wage level was W_p, employment was determined exclusively by labor demand.

The analysis in this article proceeds as follows. First the supply of labor in Tunisia is measured according to internationally comparable criteria. The criteria used in Tunisia make the labor supply curve in figure 1 appear to be much farther to the right of the diagram than in other countries. Moreover, these definitions have changed over time in a way that shifts the curve even farther to the right. Second, employment estimates for recent years are examined. The criteria used in Tunisia amount to representing the labor demand curve in figure 1 to the left of its actual location. Third, based on government placement agency records on the number of vacancies and job seekers, a rudimentary Beveridge curve is drawn. This exercise shows a movement in the direction of a tighter labor market in recent years, like the one from point A to point B in figure 2. And finally, to assess whether the labor market operates efficiently, a matching function is estimated for Tunisia. Except for first-time job seekers, the results are similar to those observed in several industrial countries for periods characterized by moderate unemployment rates. These results, together with the other evidence, suggest that unemployment is not a serious problem in Tunisia except for first-time job seekers.

Defining the Labor Force

The criteria used to measure unemployment in Tunisia differ from those commonly applied in other countries in two respects. The most frequently mentioned discrepancy concerns job seekers of ages 15 to 1[illegible], and 60 and above, who appear on the unemployment rolls in most countries but who are not included in Tunisia's unemployment figures. Because these people are available to work, observers believe that official unemployment figures underestimate the magnitude of the problem. The second discrepancy, however, more than offsets this underestimation. Tunisian data count a set of inactive people (mostly housewives) as unemployed, although they would not be counted as such elsewhere.

In 1966 the definition of the labor force used in Tunisia was similar to that used in other countries. The labor force thus included all individuals who either held a job or

were actively searching for one. In the discussion here, this definition is referred to as the *effective* labor force. In Tunisia, however, this definition was considered problematic because it yielded a higher unemployment rate for men than for women, which is uncommon. Of course, this gender gap may say more about the condition of women at that time than about the appropriateness of the definition. But in practice, the problem led to several endeavors by the statistical office (*Institut national de la statistique*) to identify potentially discouraged job seekers among inactive women. For this reason the standard definition of the economically active population was gradually extended to include the *marginal* labor force first, and then the *potential* labor force.

The marginal labor force comprises individuals who do not consider themselves economically active but who did some work in the preceding three months. More specifically, all the interviewees were first asked whether they considered themselves employed, unemployed, housewives, students, in military service, and so on. Then all the inactive respondents, apart from those in military service, were asked whether they had worked in the three months preceding the survey. Those who did were switched from inactive to marginally active, and then asked whether they worked at least one hour during the week preceding the survey. The answer was overwhelmingly negative, which is not surprising given that these individuals reported that they were inactive. As a result, most of the marginally active were counted as unemployed.

The potential labor force consists of a subset of housewives who have some free time and would like to have a job. More specifically, those interviewees who classified themselves as housewives and who did not do any other work during the previous three months, not even as family workers, were asked whether their housekeeping tasks kept them busy all day. If so, these women are considered inactive. But if not, they were asked whether they would be willing to take a job if it were offered. An affirmative answer shifts them from inactive to potentially active. Not surprisingly, most of them were reported unemployed when asked whether they worked at least one hour during the week preceding the survey.

A consistent presentation of labor force and employment data based on existing population censuses and labor force surveys is shown in table 1. In this presentation all job seekers aged 15 to 17 or 60 and above are counted as unemployed, so the data are comparable to those of other countries. The presentation also uses the questionnaires of the censuses and surveys to disentangle whether the figures include the marginal and the potential labor force. The table should be interpreted with caution though, because it is based on published data and not on the original information. In particular, when the same presentation used to construct the table is applied to much narrower groups of individuals (for example, men aged 60 and above) certain inconsistencies emerge that highlight the limits of this approach. At an aggregate level, however, table 1 yields two insights.

The first insight is that unemployment has declined over time. This is not the story line that emerges from official unemployment figures, which, after adding young

and old job seekers, have fluctuated around 15–16 percent from 1966 to 1989. (In 1980 the official unemployment rate was lower, but that year appears to be an outlier). If comparable unemployment rates are considered, instead of the official ones, the story is one of decreasing unemployment. For instance, if data on the effective labor force are used, the unemployment rate declines from 15.2 percent in 1966 to 12.1 percent in 1980 to 11.2 percent in 1989. And the trend is similar when using broader definitions of the labor force. (Again, 1980 was the only exception, a year that may be seen as idiosyncratic).

The second insight is that most of the unemployment problem is associated with new entrants to the labor force. When measured according to standard criteria, the unemployment rate was around 11.2 percent in 1989. But when first-time job seek-

Table 1. *Unemployment Rates in Tunisia*

Unemployment data	*1966*	*1975*	*1980*	*1984*	*1989*	*1990*
Number of persons (thousands)						
A Employed, ages 15+	927.3	1,366.5	1,576.9	1,786.4	1,978.8	
B Effective actives, ages 15+	1,093.7		1,793.3	-	2,229.1	—
C Marginal actives, ages 15+			16.5		89.1	—
D Effective + marginal, ages 15+		1,621.8	1,809.8		2,318.2	—
E Potential actives, ages 15+				-	42.4	—
F Total actives, ages 15+				2,137.2	2,360.6	—
G First time job seekers, ages 18–59				-	135.2	—
H First time job seekers ages 15–17				-	54.8	—
I First time job seekers, ages 15+	40.8	119.1	89.8	—	190.0	—
Official unemployment rates (percent)						
Total	15.21	15.74	12.87	16.41	16.17	—
Household heads	[illegible]		7.23			1.40
Comparable unemployment rates (percent)						
B – A)/B	15.21		12.07		11.23	—
(D – A)/D		15.74	12.87		14.64	-
(F – A)/F				16.41	16.17	—
Rates excluding first time job seekers (percent)						
(B – A – I)/(B – I)	11.92	-	7.43		2.95	—
(D – A – I)/(D – I)		9.06	8.32	—	7.02	—
(F – A – I)/(F – I)					8.84	—

— Not available

Source: *Recensement general de la population et des logements, 1966; Recensement general de la population et des logements 1975; Enquête population emploi 1980; Recensement general de la population et de l'habitat, 1984; Enquête nationale population emploi, 1989; Enquete nationale sur le budget et la consommation des ménages,* (and tables produced for the World Bank), Institut national de la statistique, Tunis.

ers are excluded, the "core" rate was less than 3 percent. Moreover, the unemploment rate of heads of households was as low as 1.4 percent in 1990. Although th(figures refer to a subset of the unemployed, they are more informative about t efficiency of the labor market than the consolidated unemployment rates. Ma first-time job seekers can be considered "voluntarily" unemployed because they ta advantage of family support to wait for the right job opening, while rejecting existi work opportunities that are attractive. Voluntary unemployment is less likely amo those who had a job and lost it.

Methods of Forecasting Unemployment

Estimates of the unemployment rate for years after 1989 are based on a comparis of employment forecasts and labor force projections. The problem here is that t method used to forecast employment leads to a downward bias, while the meth used to forecast the labor force leads to an upward bias. By underestimating t number of people at work and overestimating the number of people available f work, these two biases lead to an unrealistically high unemployment rate.

The employment forecasts are produced by the planning ministry (*Ministère développement économique*), which updates the 1989 employment figures to a count for the creation of new jobs and the destruction of old ones. New jobs a identified from employers' reports to the investment promotion agency (*Agence promotion des investissements*). Job destruction is estimated based on social securi records and on data from the labor inspection of the ministry of labor. Based this approach, 2.13 million people would have been at work in Tunisia in 199 figure used here to illustrate the bias resulting from current forecasts. But th approach has two shortcomings: the estimates of job creation and job destructi refer only to the formal sector; and they probably underestimate job creation mo sharply than job destruction.

One reason why job creation is underestimated is that the declaration of new jobs employers is not compulsory. In the past, employers had to report their investme projects and the associated new jobs to the *Agence de promotion des investissements* order to get subsidies; but reporting is now a mere formality. Moreover, employers a only asked to declare new permanent jobs, whereas employment growth in recent yea has tended to rely heavily on temporary contracts, even in the formal sector. Last bu not least, there is an evasion problem. Firms tend to declare only the personnel th register with the social security system. And they do not register all their personn because they prefer to escape payment of social security contributions.

The coverage of job destruction, which is based on social security records and dat from the labor inspection of the ministry of labor, is better than the coverage of jo creation. It still refers only to the formal sector but covers both permanent and tem

orary workers. As a result of these differences, the planning ministry underestimates employment growth in the formal sector. Moreover, the methodology used by the ministry implicitly assumes zero employment growth in the informal sector of the economy, although most observers agree that this sector has expanded since 1989. The underestimation of employment growth in both the formal and (to a much larger extent) the informal sectors suggests that employment forecasts are likely to suffer from a substantial downward bias.

Forecasts of the labor force, in turn, extrapolate from the growth rate observed in previous years. Taken at face value, the data from population censuses and labor force surveys imply the Tunisian labor force grew at a rate of roughly 3 percent a year between 1980 and 1989. Part of this growth is fictitious, however, because it results from the change in the definition of the labor force. As noted earlier, the 1989 survey included the potential labor force among the economically active, but the 1980 labor force did not. A more reliable estimate can be obtained by comparing the effective labor force in both years, which yields an annual growth rate close to 2.5 percent. The comparison between the total labor force (including marginal and potential actives) in 1984 and 1989 yields an even lower annual growth rate of about 2 percent.

A different picture of unemployment emerges depending on which of these three growth rates is used. If the 3 percent annual growth rate of the labor force is applied to the 1989 effective labor force, the corresponding figure in 1993 would have been 2.51 million. When the 2.5 percent growth rate is used instead, the effective labor force forecast is 2.46 million. At a 2 percent growth rate, it is 2.41 million. Assume for a moment that the planning ministry's employment forecast is not biased, so that total employment in 1993 can be estimated at around 2.13 million. Depending on whether the labor force actually grew at a rate of 3, 2.5, or 2 percent a year, the unemployment rate in 1993 would thus be 15.1, 13.1, or 11.7, respectively. Note that this last figure is close to the one observed in 1989, which suggests some stability of the unemployment rate in recent years. But in fact total employment is underestimated, so that the "true" unemployment rate for 1993 was likely to have been lower.

Unemployment versus Vacancies

The efficiency of the Tunisian labor market can also be assessed based on trends in the number of unfilled vacancies as well as in unemployment. For this measurement, the data collected by the government's placement agency (*Agence Tunisienne pour l'Emploi*, or ATE for short) turns out to be particularly valuable. When Tunisia was a control economy, publicly owned firms had to report all of their vacancies to the ATE, which was in charge of filling them based on its own records of job seekers. As the economy became more market oriented, the obligation of reporting vacancies to the ATE remained, but the share of hirings directly processed by firms increased.

Table 2. *Unemployment and Job Vacancies in Tunisia*

	1982	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*	*1989*	*1990*	*1991*	*1992*	*1993*	*1994*
Number of persons													
Unemployed registered	189,529	173,922	189,152	206,858	194,697	209,737	237,302	174,193	152,153	133,072	136,885	142,212	160,224
Skilled	51,931	50,292	57,496	72,379	64,940	68,215	73,482	45,554	27,363	23,960	24,035	29,184	34,692
Unskilled	82,213	67,976	73,731	80,556	76,954	77,738	84,924	60,681	65,400	57,008	53,594	57,012	57,816
First job	55,385	55,654	57,925	53,923	52,803	63,784	78,896	67,958	59,390	52,104	59,256	56,016	67,716
Vacancies reported	49,678	45,345	53,057	62,616	56,421	61,992	62,619	67,642	80,419	68,620	65,292	68,100	81,653
Skilled	18,560	18,567	23,269	25,975	21,338	24,624	26,981	29,585	34,332	29,428	31,045	34,956	44,522
Unskilled	31,118	28,778	29,788	36,641	35,083	37,368	35,638	38,057	46,087	39,192	34,247	33,144	37,131
Total placements	43,187	41,730	46,937	56,813	53,348	58,055	57,827	62,886	75,418	64,325	61,415	65,484	76,559
Skilled	12,076	12,508	15,387	19,621	17,489	19,243	19,368	20,915	16,244	15,023	17,439	18,036	20,898
Unskilled	27,313	25,655	26,125	32,259	31,861	34,508	33,104	34,425	42,239	37,805	32,379	31,848	36,845
First job	3,798	3,568	5,425	4,933	3,998	4,304	5,355	7,546	16,935	11,497	11,597	15,600	18,816
Percentage of effective labor force													
Unemployed registered	9.99	8.86	9.31	9.99	9.23	9.76	10.85	7.81	6.70	5.75	5.81	5.92	6.55
Skilled	2.74	2.56	2.83	3.50	3.08	3.18	3.36	2.04	1.21	1.04	1.02	1.22	1.42
Unskilled + first job	7.26	6.30	6.48	6.50	6.15	6.59	7.49	5.77	5.50	4.72	4.79	4.71	5.13
Vacancies reported	2.62	2.31	2.61	3.03	2.68	2.89	2.86	3.03	3.54	2.97	2.77	2.84	3.34
Skilled	0.98	0.95	1.15	1.26	1.01	1.15	1.23	1.33	1.51	1.27	1.32	1.46	1.82
Unskilled	1.64	1.47	1.47	1.77	1.66	1.74	1.63	1.71	2.03	1.69	1.45	1.38	1.52
Total placements	2.28	2.13	2.31	2.75	2.53	2.70	2.64	2.82	3.32	2.78	2.61	2.73	3.13
Skilled	0.64	0.64	0.76	0.95	0.83	0.90	0.89	0.94	0.72	0.65	0.74	0.75	0.85
Unskilled + first job	1.64	1.49	1.55	1.80	1.70	1.81	1.76	1.88	2.61	2.13	1.87	1.98	2.28

Source: Agence Tunisienne pour l'Emploi, and authors' estimates of the effective labor force

Despite the shift from a control to a market economy, however, the ATE still is an important labor market intermediary.

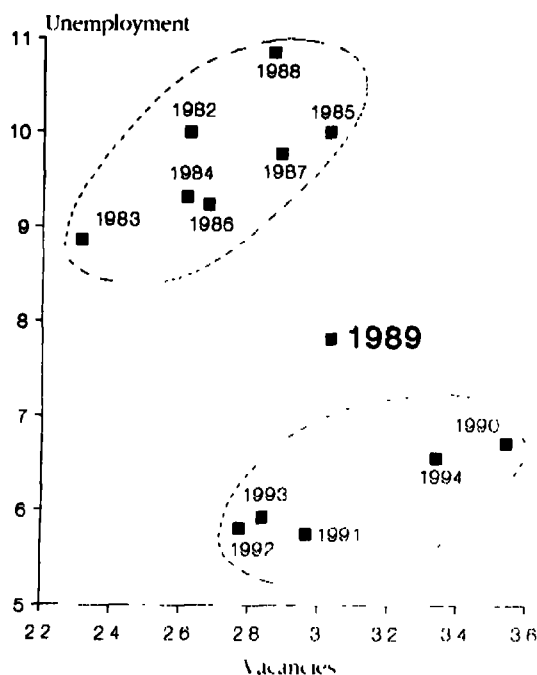
The absolute numbers of job seekers registered with, and vacancies reported to, the ATE from 1982 through 1994 is shown in table 2. The table also presents the data on job seekers and vacancies in percentage of the effective labor force, with the latter estimated based on comparable observations from population censuses and labor force surveys under the assumption of a constant growth rate between observations. The relative stability in the number of placements by the ATE is striking. If anything, the ratio between these placements and the effective labor force has gone up over time. At the beginning of the period considered, total annual placements amounted to slightly more than 2 percent of the effective labor force; at the end of the period the figure was close to 3 percent. These figures suggest that ATE records may still reflect the main trends in unemployment and vacancies, despite the shift to a market economy.

The picture that emerges from these records is one of an increasingly tight labor market. This interpretation is straightforward when the data are displayed in a Beveridge curve format, as in figure 3. Both unemployment and vacancies are measured in this figure as percentages of the effective labor force, which controls for the increasing size of the labor market. The clusters of points show that there were more vacancies and fewer job seekers in the early 1990s than in the 1980s. The similarity between the two groups highlighted in figure 3 and points A and B in figure 2 is striking and suggests that Tunisia experienced a movement along the Beveridge curve in recent years, in the direction of a tighter labor market. Given the intermediate position occupied by 1989, the year of the most recent labor force survey, it is safe to conclude that the unemployment rate must be lower now than it was at that time.

The movement along the Beveridge curve is even more pronounced when only skilled workers and vacancies are considered. At the other end of the spectrum, first-time job seekers may face higher unemployment rates despite an increase in the overall number of vacancies available. Several explanations for this trend are possible. The first is that sustained growth has created increased affluence. Families can now afford to support their young members for longer periods of unemployment, thus allowing them to find better jobs. Another possible explanation is that increased education raises the expectations of these new entrants to the labor market, making them reluctant to take the jobs that are actually available. Evaluating the empirical relevance of these two hypotheses is beyond the reach of this article, however.

Placements and the Matching Process

New jobs are created when employers find job seekers who are suitable to fill their vacancies and who are in turn interested by those vacancies. Limited job creation can

Figure 3. *Unemployment and Vacancies in Tunisia as a Percentage of the Labor Force*

Source: See table 2.

therefore be expected when few vacancies are available or when the process of matching vacancies and job seekers is very inefficient. The Beveridge curve analysis in the previous section suggests that there is no shortage of vacancies in Tunisia. According to the ATE records, many more vacancies were reported in the early 1990s than in the 1980s, even though firms are no longer obliged to report vacancies to the ATE. Therefore, if the unemployment problem were as serious as the official unemployment rates indicate, the matching process would be very inefficient in Tunisia. To assess whether this is so, matching functions estimated from ATE records can be compared to matching functions estimated in industrial countries during periods of moderate unemployment.

Whereas the ATE records distinguish between skilled and unskilled vacancies, they consider a third category of unemployed workers, namely first-time job seekers (who are not considered skilled by employers). Therefore, two separate matching functions are estimated. One explains new placements of skilled workers as a function of the number of skilled vacancies and of skilled job seekers. The other matching func-

tion explains new placements of unskilled workers as a function of the number of unskilled vacancies and of unskilled and first-time job seekers.

Matching functions are usually estimated under the assumption that a 1 percent increase in the number of vacancies leads to the same percentage change in the number of new placements, regardless of the number of job seekers. Conversely, a 1 percent increase in the number of job seekers is expected to lead to the same percentage change in the number of new placements, regardless of the number of vacancies available. In more technical terms, a Cobb-Douglas specification is used. Also, a time trend is often included in the specification. If, other things equal, the number of new placements increases over time, then it is safe to conclude that the matching process is becoming more efficient.

More specifically, matching functions usually take the form:

$$\log H_{it} = a_i + \alpha \log U_{it} + \beta \log V_{it} + \delta t + \varepsilon_{it}$$

where the subindex t corresponds to the period, and the subindex i to the district or geographic area to which the data refer. Depending on the countries, the t or the i dimension may be missing. Parameter α measures the percentage change in placements when the number of job seekers increases by 1 percent, and parameter β the corresponding change when the number of vacancies increases by 1 percent. The sum $\alpha + \beta$ indicates the returns to scale of the matching function. If this sum is close to one, doubling the number of vacancies and job seekers roughly doubles the number of new placements. Parameter δ, in turn, shows whether the efficiency of the matching process increases (for δ positive) or decreases (δ negative) over time. Finally, ε is an error term.

The results obtained when estimating matching functions on Tunisian data are reported in table 3. Note that the results probably represent a lower bound for labor

Table 3. *The Matching Function in Tunisia*

Type	Data	Constant	Time trend	Unemployment	Vacancies	Returns to scale	Adjusted R^2
Unskilled job seekers and vacancies	Yearly data 1970–94	0.684	0.0040	0.[illegible]13	0.655	0.87	0.850
		(0.5[illegible]5)	(0.832)	(3.[illegible]85)	(5.866)		
Skilled job seekers and vacancies	Yearly data, 1970–94	0.2[illegible]1	0.0261	0.014	0.930	0.94	0.912
		(0.220)	(2.899)	(0.159)	(9.669)		

Source: Author's estimates, using data from the Agence Tunisienne pour l'Emploi. The dependent variable is the log of the yearly number of placements. Independent variables are the log of registered unemployed and the log of reported vacancies. Coefficients were estimated using the AR(1) method. These coefficients are statistically significant when the corresponding t statistics (in parentheses) are higher than 2. The Durbin-Watson statistics were 2.03 for the first equation and 2.06 for the second one, which implies that there are no autocorrelation problems.

market efficiency. The data are from a government placement agency, and it would not be surprising if paperwork, red tape, and relatively low levels of effort by civil servants slowed down the placement process. Private sector matchmaking, operating through personal networks, press ads, and other informal mechanisms, is likely to be more efficient than matchmaking by the ATE. This distinction may not be that crucial when comparing the matching function in Tunisia and in industrial countries, given that the data used to estimate the latter are from public placement agencies like the ATE. Available estimates for industrial countries are reported in table A-1.

The estimates for Tunisia show a striking difference between a relatively efficient matching process for skilled workers, and a clearly inefficient one for unskilled and first-time job seekers. For the skilled segment of the labor market, these estimates are comparable to those in industrial countries. To some extent, this should not be surprising, given that a skilled job in Tunisia may be similar to an average job in an industrial country. Still, it is worth noting that the coefficients for Tunisia are almost indistinguishable from those for England and Wales, Sweden, or the United States in periods where the unemployment rate of these countries was well below 10 percent. In all of these cases, a 1 percent increase in the number of vacancies leads to larger variation in the number of placements than a 1 percent increase in the number of job seekers. In terms of the equation above, β is larger than α. Also, returns to scale (the sum $\alpha + \beta$) are close to one, or slightly below. But while the efficiency of the matching process declines over time in all of the industrial countries for which there are estimates (that is, parameter δ is negative), there is no such downward trend in the Tunisian data.

The picture is different when it comes to unskilled workers and first-time job seekers. The coefficients of the matching function in this case are not comparable to those observed in industrial countries. In particular, hirings appear to be driven by the stock of vacancies only, without any noticeable effect on the stock of job seekers. Note that this result aggregates over both first-time job seekers and unskilled unemployed, so that the role played by each of these two groups cannot be disentangled. But the lack of significance of the stock of job seekers could actually indicate that people in any or both of these two groups are not willing to take any job they are offered. This is most likely to be true in the case of first-time job seekers.

Conclusion

Tunisia's unemployment problem appears to have been greatly overstated. The use of comparable labor force definitions shows that unemployment has declined quite steadily over the years, which is hardly surprising given the remarkable performance of the Tunisian economy. The extent of the decline is partly hidden by the increase in the number of individuals seeking their first job. But the unemployment rate

excluding first-time job seekers is low by any standard, and the unemployment rate of household heads is low too. Trends in the number of vacancies relative to the number of job seekers indicate that, if anything, the unemployment rate has declined since the last available estimation. A critical evaluation of the methods used to forecast unemployment rates and an econometric analysis of the data on vacancies and unemployment also suggest a relatively efficient labor market.

Although the diagnosis in this paper is based on piecemeal evidence, the 1994 population census offers a valuable opportunity to refine it. Data from this census should be processed using the standard international definition of unemployment, in addition to the one used in Tunisia, to allow more accurate comparisons with other countries. Processing the available information on first-time job seekers would improve our understanding of the reasons for their long unemployment spells, and help in designing employment policies targeted at this specific group.

Trade liberalization and public ownership divestiture will of course entail a significant reallocation of labor across firms and activities. But it is not obvious that these reforms will lead to a surge of unemployment in the short run because first-time job seekers, who constitute the bulk of the unemployed, will be largely unaffected. The unemployment spells of first-time job seekers are usually determined by mismatch between skills and jobs, by expectations about public sector job openings, by the availability of family support during the job search, and the like. The decline and eventual increase in blue collar jobs resulting from trade liberalization and privatization is probably irrelevant in this respect. Therefore, the level of unemployment rates in Tunisia does not justify delaying much needed structural reforms.

From a broader perspective, this paper highlights some caveats that may be useful when interpreting unemployment data in developing countries, as well as some shortcuts that may help assess labor market efficiency based on these data. If piecemeal evidence can be put together in a coherent way in a country such as Tunisia, where labor market statistics are partial and scattered, it should also be feasible in other developing regions of the world, where data tend to be better on average. This kind of critical evaluation of the available statistics is particularly recommended before undertaking labor market reforms, such as changing the wage-setting mechanisms or implementing safety nets. In this respect, no analytical model of the labor market can replace a good understanding of the facts.

Table A.1. *The Matching Function in Selected Countries*

Sample and source	*Data*	*Constant*	*Time trend*	*Unem-ployment*	*Vacancies*	*Returns to scale*	*Adjusted* R^2
Austria (Christl 1992)	Monthly data, 1974–89	–1.30 (2.5)	–0.0172 (6.8)	0.76 (13.2)	0.41 (4.4)	1.17	0.66
Former Federal Republic of Germany (Burda 1993)	Daily data, by district, 1990–92			0.88 (42.7)	0.11 (9.2)	0.99	0.93
England and Wales (Coles and Smith 1996)	Data by district, for 1987	–1.19 (7.2)		0.34 (10.6)	0.66 (16.0)	1.00	[illegible]
Israel (Berman 1997)	Monthly data, 1978–90	-1.00 (–0.7)	0.0063 (3.4)	0.45 (4.9)	0.49 (5.5)	0.94	[illegible]
Sweden (Edin and Holmlund 1991)	Monthly data, 1970–88	Not reported	0.0005 (4.91)	0.23 (6.76)	0.56 (23.52)	0.79	0.91
United States (Blanchard and Diamond 1989)	Monthly data, 1968–81	0.52 (7.50)	0.0015 (2.40)	0.35 (3.90)	0.54 (6.90)	0.89	[illegible]

Note: The dependent variable is the log of placements, referrals, or flows out of unemployment. Independent variables are the log of the registered unemployed and the log of the reported vacancies. *t* statistics in parentheses. A blank indicates that the corresponding variable was not included in the econometric analysis.

Notes

Martin Rama is senior economist with the Development Research Group of the World Bank. He would like to thank Nasser Jharsalli and Ben Zarthi for helpful conversations.

1. Note that the paper does not aim at delivering a comprehensive study of the Tunisian labor market. An overall picture of the main trends in this market at different points in time is provided in Morrisson (1986) and Zouari-Bouattour (1994). Also, several papers have dealt with specific labor market issues in Tunisia, including the determinants of earnings at the individual level (Abdennadher, Karaa, and Plassard 1994), the links between poverty and employment status (World Bank 1995b), the nature of social security programs for workers (World Bank 1993), the importance of [illegible] household transfers (IRLP 1994), and the links between wage increases and productivity gains ([illegible] 1995, World Bank 1995a). The interested reader can fruitfully refer to those studies.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Abdennadher, Chokri, Adel Karaa, and Jean-Michel Plassard. 1994. "Tester la dualité du marché du travail: l'exemple de la Tunisie." *Revue d'Economie du Développement* 2(June):39–63, June.

Azam, Jean-Paul. 1995. "Labor Market Institutions." CERDI, Centre d'Études et de Recherches sur le Développement International, Clermont-Ferrand. Processed.

Berman, Eli. 1997. "Help Wanted, Job Needed: Estimates of a Matching Function from Employment Service Data." *Journal of Labor Economics* 15(1, part 2, January):S251–92.

Blanchard, Olivier, and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity* 1:1–60.

Blanchard, Olivier, and Juan F. Jimeno. 1995. "Structural Unemployment: Spain versus Portugal." *American Economic Review* 85(2) (Papers and Proceedings): 212–18.

Burda, Michael. 1993. "Modeling Exits from Unemployment in Eastern Germany: A Matching Function Approach." CEPR Discussion Paper Series 800. Centre for Economic Policy Research, London, U.K. Processed.

Christl, Josef. 1992. *The Unemployment/Vacancy Curve: Theoretical Foundation and Empirial Relevance.* New York: Springer-Verlag.

Coles, Melvyn, and Eric Smith. 1996. "Cross Section Estimation of the Matching Function: Evidence from England and Wales." *Economica* 63(252):589–97.

Edin, Per Anders, and Bertil Holmlund. 1991. "Unemployment, Vacancies, and Labour Market Programmes: Swedish Evidence." In Fiorella Padoa Schioppa, ed., *Mismatch and Labour Mobility.* Cambridge, U.K.: Cambridge University Press.

Franks, Jeffrey. 1994. "Explaining Unemployment in Spain: Structural Change, Cyclical Fluctuations, and Labor Market Rigidities." IMF Working Paper 94/102. International Monetary Fund, Washington, D.C. Processed.

Harris, John, and Michael Todaro. 1970. "Migration, Unemployment, and Development: A Two-Sector Analysis." *American Economic Review* 60(1):126–43.

Hashimoto, Masanori. 1993. "Aspects of Labor Market Adjustment in Japan." *Journal of Labor Economics* 11(1):136–61.

ILO (International Labour Office). 1996. *Yearbook of Labour Statistics.* Geneva.

INRP (Institut de recherches et d'études sur la population). 1994. *L'entraide familiale*, cahier 11. Tunis: Ministère du développement économique.

Morrisson, Christian. 1986. "Evolution of the Tunisian Labor Market." DRD Discussion Paper 279. World Bank, Development Research Department, Washington, D.C. Processed.

Pissarides, Christopher. 1985. "Short Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review* 75:676–90.

Rama, Martin. 1997. "A Labor Market Cross Country Database." World Bank, Development Research Department, Washington, D.C. Processed.

World Bank. 1993. "Republic of Tunisia: The Social Protection System." Middle East and North Africa Department, Washington, D.C. Processed.

———. 1995a. "Toward the 21st Century." Country Economic Memorandum of Tunisia. Middle East and North Africa Department, Washington, D.C. Processed.

———. 1995b. "Poverty Alleviation: Preserving Progress while Preparing for the Future." Poverty Assessment of Tunisia. Middle East and North Africa Department, Washington, D.C. Processed.

———. 1995c. *World Development Report 1995: Workers in an Integrating World.* New York: Oxford University Press.

Zouari Bouattour, Salma. 1994. "Evolution du marché du travail Tunisien, 1983–1993." Université de Sfax, Tunisia. Processed.

Evaluating Retraining Programs in OECD Countries: Lessons Learned

Amit Dar • Indermit S. Gill

Are retraining programs for the unemployed more effective than job search assistance? Governments of the member countries of the Organisation for Economic Co-operation and Development have considerable experience with retraining programs in a variety of industrial settings. Evaluations of these programs show that the results are disappointing, however. This article discusses the factors associated with retraining programs for two types of workers: those laid off en masse and the long-term unemployed. Evaluations indicate poor results for both groups: retraining programs are generally no more effective than job search assistance in increasing either reemployment probabilities or postintervention earnings, and they are between two and four times more expensive than job search assistance.

Industrial countries spend sizable amounts on labor programs for the unemployed. In 1992 member countries of the Organisation for Economic Co-operation and Development (OECD) spent between 0.1 percent (Japan) and 2.6 percent (Sweden) of gross domestic product on labor programs. In several countries this training is "the largest category of active programs, and is often perceived as the principal alternative to regular unemployment benefits" (OECD 1994a). Countries in Eastern Europe—where the role of active labor programs is a topic of current debate—also spent between 0.2 and 3 percent of gross domestic product on these programs (OECD 1994a). Yet possibly because such assistance is viewed almost as a fundamental right in Western Europe, these programs are rarely evaluated outside the United States. This article surveys evaluations of retraining programs in the OECD countries, highlighting the shortcomings of such schemes and illustrating the payoff to investing in rigorous evaluations.

We examine only the evaluations of retraining programs, so our focus is largely on adults with previous work experience rather than on unemployed school-leavers. We distinguish here between retraining schemes for workers displaced by plant closures

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 79–101

or restructuring and programs aimed at the long-term unemployed. These groups differ in important ways. First, the long-term unemployed are a relatively heterogenous group of individuals compared with those laid off en masse from a single plant or firm. Second, although some programs are targeted at specific regions, the long-term unemployed are generally more dispersed geographically. Third, the duration of unemployment is, almost by definition, greater for the long-term unemployed. Finally, retraining programs for the long-term unemployed are generally a mix of classroom and in-plant training, while programs for displaced workers are usually confined to the classroom.

Evaluation Techniques

Techniques for evaluating the effectiveness of retraining programs can be broadly classified into two categories: scientific and nonscientific. Scientific evaluations can be further divided into experimental and quasi-experimental. Experimental, or classically designed, evaluations require selecting "treatment" and "control" groups before the intervention: the treatment group receives the assistance, and the control group does not. If large numbers of individuals are randomly assigned to each group the average characteristics of the two groups should not differ significantly, so any difference in outcomes can be attributed to program participation. In quasi-experimental studies treatment and control groups are selected after the intervention. To compute the program's effects, statistical techniques are used to correct for differences in characteristics between the two groups. Nonscientific techniques do not use control groups to evaluate the effect of interventions but instead rely on statistics compiled by a program's administrators. These evaluations are of little use without a control group, it is difficult to attribute the success or failure of the participants to the intervention, because the changes in individuals' behavior might have resulted from other factors, such as worker-specific attributes or economywide changes (Grossman 1994).

Classically Designed (Randomized) Experiments

This technique, which was originally developed to test drug effectiveness, identifies and randomly assigns individuals to either a treatment or a control group before the intervention. Its main appeal lies in the simplicity of interpreting results—the effectiveness of the program is computed as the simple difference in the outcome between the participants in the treatment group and the nonparticipants in the control group (Newman, Rawlings, and Gertler 1994). The main pitfalls of this method are a failure to select individuals through random assignment, changes in behavior as a result of their assignment to either group (for instance, enrolling in private programs or intensi-

fying their job search), high costs because of the number of participants in the sample, and ethical questions about excluding a group of people from the intervention.

Although randomization is thought to eliminate selection bias in determining who will participate, proponents of this methodology make an important assumption: that random assignment does not alter the behavior that is being studied. This may not be the case and, in fact, the bias may be quite strong (Heckman 1992). For example, ambitious individuals who would have taken a training course in any case will not apply to the program if they stand a chance of being in the nontreatment group. Such individuals, who might have enrolled in a nonrandomized regime, may make plans anticipating enrollment in a training program. With randomization they may alter their decision to apply or undertake activities complementary to training. Thus risk-averse persons will tend to be eliminated from the program.

Quasi-Experimental Techniques

In quasi-experimental methods the treatment and control groups are selected *after* the intervention. Econometric techniques are used to correct for the differences in characteristics between the two groups. The main appeal of this method lies in its relatively low cost and the ability to undertake the evaluation at any time. The main drawback is that these techniques—if done properly—are statistically complex. The techniques used to adjust for differences in *observable* attributes (for example, sex, age, region of residence, education) are relatively straightforward but subject to specification errors; correcting for *unobservable* characteristics (for example, motivation, family connections) requires a convoluted procedure that can yield wildly different results. Quasi-experimental evaluations fall into three types: regression adjusted for observable characteristics, regression adjusted for observable and unobservable characteristics (selectivity corrected); and matched pairs.

REGRESSION ADJUSTED FOR OBSERVED CHARACTERISTICS. This technique is used to assess the impact of participation in a program when the observable characteristics of the participants and the comparison group are different. This method is appropriate for estimating the effect of a program when the difference between participants and nonparticipants can be explained by these observable characteristics. For example, if better educated workers are more successful in finding work regardless of whether or not they had special training, then controlling for the effect of education (using regression techniques) will provide more reliable estimates than would a simple comparison of the reemployment probabilities of the control and treatment groups.

REGRESSION ADJUSTED FOR OBSERVED AND UNOBSERVED VARIABLES (SELECTIVITY CORRECTED). When selection into programs is not random, and participation in a program is due to both observable and unobservable characteristics, the above tech-

nique, which corrects for observed characteristics, is likely to be biased. Even if participants and nonparticipants have similar observable characteristics, unobservable characteristics (such as innate ability) would lead to incorrect inferences about nonparticipants. This technique uses a method developed by Heckman (1979) called "sample selectivity correction" to try to control for these unobservables.

MATCHED PAIRS. Because the observed characteristics of individuals in the control and treatment groups are bound to be different to some degree, these groups are likely to have different success rates in finding employment even in the absence of active labor market programs. To control for these differences, synthetic control groups are constructed using a matched pairs approach. The synthetic control group which is a subset of the entire control group, is composed of individuals whose observable characteristics most closely match those of the treatment group.

Relative Strengths of Techniques

Estimating the effect of an employment program on the earnings of trainees, using randomized and quasi-experimental techniques, LaLonde (1986) found that randomized experiments yielded results significantly different from those that relied on quasi-experimental techniques. Policymakers should be aware that available nonexperimental evaluations of training programs may contain large biases. While randomized experimentation is theoretically the best technique to estimate the effects of interventions, quasi-experimental techniques may be superior in practice.

The main weakness of randomized experiments is their inability to ensure that individuals in the control group do not alter their behavior in a way that contaminates the experiment. For example, individuals denied public job training might enroll in private programs, which would bias the results of any evaluation of public programs.[1] It may also be difficult to ensure that assignment is truly random. For example, applicants may be selected into the program because of nepotism, or program administrators may deliberately exclude high-risk applicants to achieve results that reflect well on the program. A third problem concerns ethical questions about treating individuals as subjects in an experiment. Finally, experimental evaluations are possible only for future programs, because the control and treatment groups have to be selected before the program is initiated.

Using the dual criteria of rigor and feasibility then, randomized experiments are not necessarily superior to quasi-experimental techniques. Because the decision to evaluate labor market programs often occurs after the programs are in place and because the costs of setting up the experiments are high, randomized evaluations should perhaps be the last alternative. Within quasi-experimental techniques, selectivity correction may not add much, especially when information is available for a considerable number of observable individual and labor-market characteristics (such

as education, age, sex, household wealth, and region of residence). In addition to being cumbersome and somewhat counterintuitive, this method often gives arbitrary results depending on the selectivity-correction specification that is used.

This leaves the matched pairs and regression-adjusted techniques. Of the two, the matched pairs technique is preferred because the procedure is less arbitrary. Because the observed differences between the treatment and comparison group are minimized, the exact specification of the model becomes less important. And because the program measures the simple difference in the variables that policymakers want answered (reemployment probabilities and wages) between the control and treatment groups, the results are easier for nonstatisticians to interpret.

One weakness shared by both the scientific and nonscientific evaluations is that they do not take into account the displacement that may result from the retraining program. For example, in countries where demand for labor is constrained, retrainees may "bump," or displace, employed workers, so aggregate unemployment may not change despite the size of the program. In general, displacement implies that the social benefits from reemployment attributable to the retraining program are lower than indicated by the evaluation, however well done.

The Importance of Costs

For the purposes of informing policy decisions, an evaluation is not complete until the costs of both the retraining program and its alternatives are considered. For example, if retraining is twice as costly as job search assistance but no more effective in finding people jobs and increasing their wages, job search assistance is twice as cost-effective. At least at the margin, such a finding would constitute a case for reallocating resources from retraining to job search programs. Unfortunately, costs appear to be the least analyzed aspect of these programs in OECD countries.

Even the most careful evaluations of retraining programs cannot be used for *social* cost-benefit analysis because of the displacement effects discussed above. But when done correctly, evaluations are good guides for cost-benefit analysis of *private* training programs, which policymakers can use to institute cost recovery in public programs and to promote private provision. Evaluations may also help in deciding whether retraining programs reduce budgetary expenditures by moving people off unemployment benefits into productive employment or whether the programs are a drain on the budget despite being effective.

Evaluating Retraining in OECD Countries

Retraining programs in OECD countries have been designed primarily to assist three categories of workers: those laid off en masse; those who have lost their jobs because

of plant closures; and those who have been unemployed over the long term. This review of the evaluations of eleven retraining programs classifies the specific situation that the program was designed to address, describes the policy intervention that was selected and the type of evaluation used, and reports the main results. Because the success of retraining programs depends on aggregate or regional labor market conditions, such as unemployment rates and the state of the leading industry, these indicators are reported as well.

We examine the type of retraining provided, whether in classrooms or on-the-job and whether it was accompanied by—or in lieu of—other measures, such as job search assistance. The evaluations are classified as experimental, quasi-experimental, and nonscientific, and we do not consider inferences drawn from the nonscientific evaluations to be reliable. We then look at the effects of the program on reemployment and wages, both for subgroups of retrainees and type of intervention. The costs of the program are included when they are reported.

Training Programs Instituted as a Result of Mass Layoffs and Plant Closures

The results of eleven programs (three in the United States, four in Sweden, and one each in Australia, Canada, Denmark, and France) examining the effectiveness of retraining programs for workers displaced through mass layoffs and plant closures have been reviewed (Table A.1). Five of the evaluations were nonscientific and five were quasi-experimental. One study relied on more than one technique to evaluate the impact of the program; no study used experimental evaluation techniques.

The retraining programs were undertaken to assist workers in the steel, pulp mining, shipbuilding, and automotive industries. The number of workers who lost their jobs varied from about 500 to 3,000 per plant. The rationale underlying the programs was apparently to assist the affected workers in any way possible. Generally these programs were instituted during periods of high or rising aggregate unemployment or during a contraction in certain manufacturing industries. For example the evaluations in the United States and Canada covered primarily the auto and steel industries, which were battered by competition from Japan. Between 1978 and 1980 auto production in the two countries declined 25 percent, precipitating layoffs and plant closures in the early 1980s. In Europe and Australia the retraining programs seem to have been instituted during periods of high or rising rates of unemployment. Most of the retraining programs were classroom based, and accompanied by job search assistance. With only one exception, on-the-job training was not provided or facilitated. In France retraining was accompanied by financial incentives to regional firms that hired trainees, so the full costs of the program were likely high (OECD 1993a).

Quasi-experimental and nonscientific techniques were used in evaluating all of these retraining programs. None of the studies was longitudinal, so they provide at best a snapshot of the labor market benefits of the program. The longer-term benefits of retraining were not evaluated even by the scientific evaluations. Although nonscientific evaluations indicate that these programs are very effective in placing high numbers of male workers in wage-earning jobs or in self-employment and women in business start-ups, more reliable quasi-experimental evaluations indicate that retraining programs are generally no more effective than job search assistance in increasing either reemployment probabilities or postintervention earnings. Some retraining programs resulted in modest gains in reemployment probabilities, but the wage changes were sometimes negative. Interestingly, evaluations of three retraining programs for U.S. auto workers showed the contrast between scientific and nonscientific techniques: in San José, California, a nonscientific evaluation indicated high placement rates, while in Buffalo, New York, and Michigan—during the same period—scientific evaluations showed that these programs were ineffective (OECD 1993a; Corson, Long, and Maynard 1985; Leigh 1992).

The costs, when they are known, varied between $3,500 and $25,000 a person. Evaluations seldom report the full costs of retraining or job search programs, however, so determining cost-effectiveness is difficult. Retraining programs appear to be between two and four times more expensive than job search programs: for example, in Buffalo, job search services cost $850 a participant, while retraining cost $3,300 (Corson, Long, and Maynard 1985). If, as the findings indicate, both programs have roughly the same success, job search assistance may be more cost-effective than retraining in assisting displaced workers get jobs.

Training Programs for the Long-Term Unemployed

There is no reason to assume that the impact of retraining on the long-term unemployed is the same as it is on workers laid off en masse. The results of studies examining the effectiveness of retraining programs for the long-term unemployed are shown in Table A.2. Of the eleven evaluations (four in the United States, three in Germany, two in the Netherlands, and one each in Canada and Britain), four were nonscientific, four were quasi-experimental, and three were experimental.

The clientele of retraining programs for the long-term unemployed is relatively heterogeneous. Because these individuals are displaced from various sectors and some have never worked, they are likely to be more varied in age, skills, and education than laid-off workers. New Jersey's retraining program in 1986–87 included workers whose previous jobs were in manufacturing, trade, and services (Anderson, Corson, and Decker 1991); many were more than 55 years old. In contrast, half the workers in Germany's retraining program for the long-term unemployed were less than 35 years old (Johanson 1994).

These programs are generally instituted during improving conditions in industry or in aggregate employment. Interventions at this stage, if appropriately designed, are thought to enable the long-term unemployed to obtain some of the jobs that are being created. The programs are more comprehensive than are those for people laid off en masse and generally provide a mix of classroom or workshop training, on-the-job training, and job search assistance.

Some experimental evaluations were conducted in the United States, and quasi experimental techniques were used in both the United States and Europe. Some of these studies were longitudinal, providing an indication of the medium-term labor market benefits of the retraining program (see Johanson 1994 and OECD 1993b for Europe; and Leigh 1992, and Anderson, Corson, and Decker 1991 for the United States). But many of the evaluations in Europe were nonscientific.

Again, the results of nonscientific evaluations, which were encouraging, were not confirmed by evaluations based on scientific techniques. The effectiveness of programs for the long-term unemployed, while not high, was better than that of programs for those laid off en masse. A few programs did result, or were thought to have resulted, in gains in either reemployment probabilities or wages; some evaluations also indicated that these programs were more effective in helping women. But where participants did record gains in employment, longitudinal studies generally found that the effects of retraining dissipated within a couple of years. In this group, too, retraining programs were generally no more effective than job search assistance in increasing either reemployment probabilities or postintervention earnings. For example, evaluations of the Texas Worker Adjustment Demonstration program indicated that participants were likely to be employed more rapidly than nonparticipants, but that over time, the employment opportunities of male participants were no better than those of nonparticipants or of those who only had job search assistance (Bloom 1990). The costs, when known, varied between $900 and $12,000 a person, about twice as much as job search services, but the lack of data makes it difficult to determine the absolute cost-effectiveness of these programs. Still, job search assistance programs appear to be somewhat more cost-effective than retraining programs in finding jobs for the long-term unemployed.

Evaluation Results for Hungary

Rising unemployment and falling real wages are a vivid and costly aspect of countries in transition from controlled to market economies. Long-run unemployment is particularly pernicious, and many countries have mounted active labor programs (for instance, public retraining programs and public service employment) to deal with this problem. With the exception of Hungary, where the World Bank has sponsored rigorous evaluations and collections of data on costs, little reliable information exists

n their effectiveness. As an early reformer, Hungary provides valuable lessons for ther transition economies. Since 1989 sharp declines in the country's gross domes-c product have been accompanied by rising unemployment and falling real wages. he unemployment rate appears to have stabilized since 1993, but employment ontinues to decline, reflecting a continuing withdrawal of workers from the formal bor market. Since 1990 the government has offered workers retraining programs nanced by a national employment fund. Trainees, who are either currently unem-loyed or are working but expect to become unemployed, include participants in ublic works programs. Instruction is largely classroom based.

Evaluations using different quasi-experimental techniques—matched pairs, regres-on adjusted, and selectivity corrected—yielded different results (O'Leary 1995). efore adjusting for differences in the characteristics of program participants and ontrol groups, estimates of effectiveness indicated that retraining significantly raised e probability of reemployment. But when the observable characteristics of treat-ent and control groups were taken into account, retraining (and training, because bout 40 percent of trainees had not worked previously) was only marginally suc-essful at best, increasing the probability of finding employment by 6 percent. Fur-her controls for unobservable attributes led to ambiguous results. Retraining was ot at all successful in raising earnings.

Preliminary analysis indicates that retraining was a substitute for attributes that ad to higher reemployment probabilities in the absence of any intervention, such as eing younger, better educated, and from more dynamic regions. That is, the rogram's value-added (in terms of improving labor market outcomes) is greater for latively disadvantaged job-seekers. This finding may be country specific, and other ountries or regions should determine whether it applies to them before implement-ng similar large scale programs. For the programs evaluated in Hungary, focusing ore on job seekers who lack these attributes would appear to serve both equity and fficiency objectives better than simply ensuring support for programs whose re-rainees have high probabilities of reemployment and gains in earnings. The analysis mplies that public retraining programs should target older men from backward egions.

This finding also highlights the usefulness of rigorous impact evaluations, which et out the effects of such attributes in determining whether a program is effective. And it underscores the need to agree upon a reliable, feasible, and easily interpreted echnique to evaluate the efficiency and equity effects of labor programs. Because of ts analytical rigor and feasibility, we argue that the preferred evaluation technique is matched pairs analysis, where trainees are compared with a subset of the control roup whose characteristics most resemble their own.

Private cost-benefit analysis of the costs of retraining (an average of approximately 900 per trainee in 1994, according to Pulay 1995) and the level of gains in reem-ployment reveal that based on reasonable assumptions about the durability of

Table 1. *Summary of OECD Retraining Programs and Evaluations*

Item	*Displaced workers*	*Long-term unemployed*
Previous sector of employment	Manufacturing	Heterogeneous
Labor market conditions	Generally deteriorating	Generally improving
Training venue	Mainly classroom	Classroom and OJT
Other measures provided	JSA	JSA
Type of evaluation		
Nonscientific	6	4
Quasi-experimental	5	4
Experimental	0	3 (all U.S.)
Effectiveness		
Nonscientific techniques	Positive	Generally positive
Scientific techniques	Negative	Generally negative; some groups benefit
Relative to JSA	No more effective	No more effective
Costs		
Retraining	Data are not available	Data are not available
Relative to JSA	At least twice as costly	At least twice as costly

Note: JSA, job search assistance. OJT, on-the-job training.

program effects and the amount and duration of unemployment benefits, it would take more than 30 years to recover costs of the program (Gill and Dar 1995). It seems difficult to justify retraining programs based on economic considerations alone, even in a country such as Hungary where the government provides benefits for unemployed workers.

Summary and Conclusion

The paucity of rigorous evidence on the costs and effectiveness of retraining programs does not permit a definitive conclusion on whether such interventions can be justified economically (Table 1). The scattered evidence does not appear to justify the indiscriminate expansion of retraining programs to cover more of the unemployed. These conclusions are consistent with the findings in the *OECD Employment Outlook* (1993b), which concluded that "for the broadly targeted sub-group of programs, the overall impression is most troubling. Available evidence does not permit strong conclusions, but it gives remarkably meager support of a hypothesis that such programs are effective."

From this discussion it appears that transition and other restructuring economies can draw the following lessons from experience in the OECD countries:

First, sound techniques should be used to evaluate retraining programs (and other public interventions). Although the nonscientific evaluations of retraining programs present a rosy picture based on placement rates and other informal evidence, scientific evaluations are quite discouraging. Relying on nonscientific evaluations may lead to incorrect policy conclusions.

Second, rigorous evaluations, although not necessarily allowing a complete social cost-benefit analysis, can be useful for policymakers in allocating public expenditure on labor programs. Reviews of evaluations find, for example, that job search assistance measures —which cost less than retraining but appear equally effective—may be a more cost-effective device in assisting displaced workers.

Third, OECD experience of retraining programs for workers displaced en masse may be useful in designing assistance programs in transition countries such as the former Soviet Union and liberalizing economies that expect labor shedding in the manufacturing and mining sectors. Principally, these economies should recognize that retraining should not be the main form of assistance.

Fourth, transition and developing economies that are beginning to experience long term unemployment can learn from OECD experience, which indicates that retraining programs for the long term unemployed are more beneficial for some groups than for others within this relatively heterogeneous group of job-seekers. Which group will benefit most from retraining is difficult to predict, however. Principally, these results call for using modest untargeted pilot programs, evaluating them rigorously, and then tightly targeting public retraining programs to those for whom they are found most cost effective.

Notes

Dar is in the Social Protection Group of the World Bank's Human Development Network, and Indermit S. Gill is senior economist in the World Bank's Country Management Unit in Brasilia. The authors are grateful to Jane Armitage, Kathie Krumm, Ana Revenga, Michal Rutkowski, and anonymous referees for helpful suggestions.

Heckman (1992) documents other limitations of this technique when applied to social experiments; these limitations derive from selectivity biases.

These findings contrast with those of Revenga, Riboud, and Tan (1994) for retraining programs in Mexico. Using quasi-experimental techniques, they find that effectiveness can be improved if programs better target relatively educated job seekers of both sexes with previous work experience. They also find generally more encouraging results than those for OECD countries and Hungary.

Table A.1. *Overview of Studies Evaluating Training Programs for Workers Displaced by Plant Closures and Mass Layoffs*

Labor market problem	*Relevant indicators*	*Intervention design*	*Type of evaluation*	*Result*	*Comments*
Ford Motor Co. plant closure in San José, Calif. (1982). Some 2,400 workers affected (OECD 1993a).	A 25 percent decline in auto production between 1978 and 1980. Unemployment rates rose from 7.5 percent in 1981 to 9.5 percent in 1982 and 1983, and manufacturing employment declined by 5 percent during 1981–83.	Basic skills training as well as targeted vocational training in marketable skills.	Nonscientific	High rate of success in placing workers in jobs.	Perceived success was due to adequate resource base ($6,000 grant per worker) and a high degree of coordination provided by Ford and by government.
Automotive plant closures in Michigan (1980–83). More than 3,000 workers laid off, mostly experienced, highly paid blue-collar male employees (Leigh 1992; 1994).	A 25 percent decline in auto production between 1978 and 1980. High unemployment rates in 1981–82 nationally (9.5 percent), which fell to 7.5 percent by 1984. Manufacturing employment rose by 5 percent between 1983 and 1984	Both JSA and classroom training were provided promptly after plant closures. Retraining was offered in occupations in which there was growing demand, including blue-collar trades	Quasi-experimental	Classroom training did not significantly improve the participants' reemployment rate. Trainees did no better than those receiving JSA.	Earnings estimates vary (ranging from negative to significantly positive). Training does not seem to have been very effective, especially in light of the fact that training cost twice as much as JSA.
Auto and steel plant closures in Buffalo, N.Y. (1982–83). Mass layoffs of highly paid, experienced blue-collar male workers ([illegible]	High unemployment rates in 1981–82 nationally (9.5 percent), which fell to [illegible] percent by 198[illegible]. Manufacturing	Displaced workers were provided with both JSA and either classroom training or OJT. Program services were provided [illegible]	Quasi-experimental	JSA services had a fairly large impact on earnings for the first six post-program months. However, there is no evidence that either	Classroom training and OJT were ineffective. These programs cost about four times as much as JSA, implying that JSA is *potentially*

[illegible]	[illegible] between [illegible] and [illegible]	[illegible] unemployment [illegible]		[illegible] had any incremental effect.	the only cost effective program. No evidence was provided on the participants' employability.
Plant closures at 13 steel factories and mines in Canada in the 1980s (Leigh 1992).	Unemployment rates rose between 1980 and 1983. Employment in nonagricultural activities fell slightly during the period.	JSA and training programs were provided.	Quasi-experimental	Participants had a 7 percent higher rate of employment than comparably displaced workers who did not participate in the program. This impact was attributed to training in which 28 percent of workers participated. But at two mining sites, the program had no impact.	Training seems to have had a greater impact than JSA. No information was available on the cost of training. Job counseling has little impact in Canada, where such services are routinely provided to the unemployed by the Public Employment Service.
Steel and coal plant closures in Creusot-Loire, France (1984, 1,230 people affected) (OECD 1993a).	Steel sector contraction resulted in the loss of 6,000 jobs during 1984. Unemployment rates rose from 8.1 percent in 1982 to 10.2 percent by 1985. In mining employment fell by almost 40 percent between 1980 and 1990. Manufacturing employment fell about 6 percent from 1983 to 1985.	Workers received 70 percent of their former salaries for 10 months while enrolled in retraining and job-search activities. Trainees (in engineering, plastic molding, refrigeration, etc.) were promised reemployment in the region. Career counseling and short courses in work skills, production methods for small firms,	Nonscientific	High success rate in finding jobs for participants.	No evidence that retraining programs produced long-term employment benefits.

(Table continues on the following pages.)

Table A.1. (continued)

Labor market problem	Relevant indicators	Intervention design	Type of evaluation	Result	Comments
		and job search techniques were also provided. Firms were given financial incentives to hire these workers.			
Pulp plant closure in Kramfors, Sweden (1977) (OECD 1991).	Unemployment rates in Sweden rose from 1.6 percent in 1976 to 2.2 percent by 1978. Employment in manufacturing declined by 8 percent between 1977 and 1979.	Participants were provided with classroom training.	Quasi-experimental	Participants received lower weekly wages than those who did not receive training. Drop in earnings was especially significant in the first year; there were no appreciable long-term gains.	No information provided on reemployment rates or costs. Benefits from retraining program were insignificant.
About 2,000 workers laid off at the Uddevalla shipyard in Sweden (1985) (Alfthan and Janzon 1994).	Volvo Co. set up a manufacturing plant in Uddevalla. Unemployment rates were declining in this period while manufacturing employment was fairly steady.	A significant number of workers joined retraining programs several months before they were laid off. Courses of varying duration were offered in welding, engineering, and control engineering. These retraining programs were provided by the state-owned training board, municipal education institutions and other adult education services.	Nonscientific	By November 1987, more than 90 percent of the workers who had completed training had found jobs or become self-employed—most of them in occupations for which they had trained.	Two factors accounted for the success of the training program: economic and labor market conditions were buoyant in the region, and management, employment offices, and training agencies worked in close cooperation.

Volvo Co. plant at Goteborg, Sweden, announced that 1,000 workers would be laid off in 1992 as it phased out an old model and tooled up for a new one (Alfthan and Janzon 1994).	Volvo planned to recruit 800 workers to manufacture the new line of cars. Unemployment rates were rising sharply—from 3.2 percent in 1991 to 5.9 percent in 1992. Manufacturing employment dropped by 9 percent in 1992. Total employment fell by 4 percent.	A retraining program was designed to help the existing work force manage the change without job losses. The company accepted the proposal under the stipulation that the cost be shared by the government. The program was a broad competency-raising exercise, with specific training to prepare the participants for the production of the new automobiles.	Nonscientific	Program has not yet been evaluated.	Program costs are expected to be about $25 million ($25,000 per person), close to half of which will be paid for by the government. In judging the cost-effectiveness of this program, it should be noted that the unemployment benefits the government would have had to pay would amount to $6.5 million.
Shipyard closure in Storstrom county, Denmark (1986). 2,000 people lost their jobs (OECD 1993a).	High unemployment rates in region, especially among women. National unemployment rates for men in 1986 and 1987 were 6.1 and 6.4 percent, respectively, while the corresponding numbers for women were 10 and 9.6 percent. Employment in manufacturing was stagnant from the mid-1980s.	This program (1986–89) trained women entrepreneurs to help them start their own businesses. A total of 200 hours of specific business-oriented training was provided.	Nonscientific	Fifty-one businesses were set up in 1989. Less than a third of the participants opened a full-time business, and few hired any employees, thus the total additional employment generated was low.	The program did not seem to be successful and was apparently quite costly (reliable cost estimates are not available).

(Table continues on the following page.)

Table A.1. *(continued)*

Labor market problem	*Relevant indicators*	*Intervention design*	*Type of evaluation*	*Result*	*Comments*
Sweden, 1980s and 1990s: general evaluation of public retraining programs (OECD 1991).	Unemployment rose steadily over the period.	Various types	Various types	Retraining programs have become less effective over time—especially since the economy has begun deteriorating. Participants have had more trouble finding jobs than other unemployed workers.	The cost-effectiveness of these programs declined both because they were less successful and because the costs increased.
Automobile plant closure in Australia in 1984. 445 workers laid off (Leigh 1992).	Unemployment rates rose steadily in Australia, reaching a high of 9 percent in 1984 before falling to 8 percent for the next few years. Between 1980 and 1984 employment in manufacturing fell by 4 percent	Labor Adjustment Training Arrangement offered classroom training (average length 19 weeks). Main distinction between courses was whether they provided driver training or not.	Quasi-experimental	Over a nine-month period, driver training increased the probability of reemployment. However, fewer trainees of *other training courses* found jobs.	There were no data showing the cost-effectiveness of different types of training courses. Self-selection was a problem because individuals who chose not to participate were in the control group. The impact of longer training was negative.

Note: JSA, job search assistance, OJT on-the-job training.

Table A.2 Overview of Studies Evaluating Training Programs for the Long-Term Unemployed

Labor market problem	Relevant indicators	Intervention design	Type of evaluation	Result	Comments
Training for workers eligible for JTPA Title III program (displaced worker) in Houston, Texas (1983–85). Eligible individuals were unemployed and had a low probability of returning to their previous occupation or industry (Bloom 1990).	A decline in petrochemical industry led to layoff. U.S. unemployment rates declined from 9.5 percent in 1983 to 7.4 percent in 1985 (and maintained this trend until the late 1980s). Employment in this industry increased slightly between 1983 and 1985.	Texas WAD project. Displaced workers were provided with JSA or a mix of JSA with classroom training.	Experimental	One year after the program's completion, participants registered no additional earning gain when compared with the sample of workers who had been given only JSA.	Despite the high costs of classroom training (twice as much as JSA), no additional gains accrued from this type of training.
Training for workers eligible for JTPA Title III program (displaced worker) in El Paso, Texas (1983–85). (Criteria same as above.) Workers laid off from light manufacturing plants (Bloom 1990).	U.S. unemployment rates declined from 9.5 percent in 1983 to 7.4 percent in 1985 (and maintained this trend until the late 1980s). Employment in manufacturing fell by close to 2 percent between 1982 and 1983 but rose to about 5 percent in 1985.	Texas WAD project. Displaced workers provided with JSA alone or combined with classroom training.	Experimental	One year after the program's completion, this program recorded no effect on male employees' earnings but did increase the earnings of women.	The increase in earnings of female participants was slightly more than the program's costs. Male employees reported no beneficial effects.

(Table continues on the following pages.)

Table A.2. *(continued)*

Labor market problem	*Relevant indicators*	*Intervention design*	*Type of evaluation*	*Result*	*Comments*
Assistance aimed at the long-term unemployed in New Jersey (1986–87). In general, workers were laid off from manufacturing, trade, and services (Corson and others 1989).	Unemployment rates in the U.S. fell from 7.1 percent in 1985 to 6.1 percent by 1987. While employment in manufacturing remained fairly steady between 1985 and 1988, trade employment increased 7 percent and service-sector employment rose about 5 percent.	New Jersey UI Reemployment Demonstration project. This program provided displaced workers with JSA alone, or followed by OJT, classroom training (CT), or a reemployment bonus.	Experimental	More than 10 quarters after the program ended, both CT and OJT significantly increased earnings over those who received JSA only. These individuals were also employed for longer periods of time in each quarter than the JSA-only group.	While the results of training were positive, the following caveats apply: the trainees were self-selected (thus these results may not apply for a random group of claimants); only 15 percent of those offered training accepted it; and training benefited those who already had marketable skills. Finally, analysis showed that the costs exceeded the expected benefits.
Long-term unemployed workers in the U.S. (1988), primarily resulting from plant closures (Leigh 1992).	Unemployment rates in the U.S. fell from 7.1 percent in 1985 to 6.1 percent by 1987. While employment in manufacturing remained fairly steady between 1985 and 1988, trade employment increased 7 percent and that in ser-	The Trade Adjustment Assistance Program was intended to develop workers' skills in new occupations. Most of these skills were supplied by a vocational college or local community college in courses that were more than a year long	Quasi-experimental	Individuals who received training began earning significantly more than those who received extended income-maintenance benefits by the 6th quarter and continued to earn more until the 12th (last) quarter, reaching a level of [illegible]	The analysis was limited to the manufacturing sector. Long-term investments in training may be effective in increasing earnings, but the training is costly (each trainee was given a $12,000 training voucher).

Long-term unemployed workers in Canada (Goss and associates 1989)	Compared with the U.S. unemployment rates have been fairly high in Canada. However they declined from 11.2 percent in the mid-1980s to 7.5 percent in 1989. Spurred by growth in commerce and services, employment grew about 3 percent annually over the time period	The Job Development Program provided classroom training and OJT. Wages and direct costs of classroom training were subsidized. The wage subsidy helped employers cover the cost of OJT	Quasi experimental	Employability of women rose while that of men declined. Weekly earnings of women were insignificant relative to that of the control group, while they were lower for males.	In view of the high program costs (around $9,300 per participant), this training is not cost-effective, especially for men.
High regional unemployment levels in Germany in the late 1970s (OECD 1991)	In 23 of the 142 regions in Germany, unemployment levels were above 6 percent. Some firms were also facing serious problems in adjusting to economic changes.	Among other interventions, training and retraining of unemployed in firms. The firms who employed these trainees received a subsidy of 90 percent of the employees' wages for two years.	Nonscientific	Training reduced unemployment somewhat, but it is estimated that more than 40 percent of these hard-to-place individuals had already left their jobs by 1981.	In light of the extremely high cost (around $500 million for training and other interventions), the results were very disappointing. No information was provided on wages.
Individuals at risk of being unemployed in 1987–88 in Germany (Johanson 1994).	Unemployment rates were steady between 1986 and 1988. They fell somewhat by 1990. Employment grew at slightly more than 1 percent during this period.	This is an evaluation of four programs, two of which trained workers for new occupations (one offered further training for employed and unemployed individuals, while the other offered retraining for the unemployed).	Quasi-experimental	No type of training had any significant impact on the flows out of short- or long-term unemployment, or on the flows into unemployment.	No information on wages or costs of training.

(Table continues on the following pages.)

Table A.2. *(continued)*

Labor market problem	*Relevant indicators*	*Intervention design*	*Type of evaluation*	*Result*	*Comments*
Individuals who were either at risk of becoming unemployed or were already unemployed in Germany in the early 1980s (Johanson 1994).	Unemployment rates remained fairly steady at just over 10 percent in the early 1980s.	Training prepared workers for certification in one of the 375 apprenticeable trades The program required up to two years to complete. Contents and specification of retraining correspond to those of initial vocational training.	Nonscientific	The overall success rate, measured in terms of retention rates, was 54 percent. Employability varied by age—less than half of those over 45 years old who were unemployed more than one year found jobs, while 86 percent of 25- to 35-year-olds did.	No data on cost were available. The number of workers who left their jobs was fairly high: two years after completing training only 60 percent of the men and 66 percent of the women were still employed.
Help for long-term unemployed and displaced workers in England in the 1980s (Addison and Siebert 1994).	High unemployment rates (around 10–11 percent) persisted through much of the 1980s. Between 1983 and 1990 employment rose by around 1.5 percent	This community-based program provided vocational training by local authorities in conjunction with local colleges, central government, and voluntary organizations. Little private sector involvement	Nonscientific	Little impact on flows from long-term unemployment.	The cost-effectiveness of these programs is likely to be negative.

Improve skills and provide better access to jobs for the long-term unemployed in Tilburg, the Netherlands (OECD 1993a).	Unemployment rates fell from about 13 percent in the early 1980s to below 10 percent by the end of the decade	Training provided hands-on experience (*through a simulated workshop*). The emphasis was on technical skills (metalworking, woodworking, installation techniques, and apparel trades) as well as on good work habits. Courses ran from 4 to 10 months.	Nonscientific	In 1991 of the 82 individuals enrolled in the course, 52 completed it and 41 found jobs or went on for further education. The most successful training was in metalworking.	Training was quite expensive: annual funding of the program was about $7.5 million, or $10,000 per trainee.
Job training programs for unemployed and displaced workers in the Netherlands (OECD 1994b).	Unemployment rates fell from about 13 percent in the early 1980s to below 10 percent by the end of the decade.	Centers for adult vocational training.	Quasi-experimental	Unemployed people who were not in the program found jobs as quickly as those who were. Even two years later, the employment situation for the two groups was not significantly different.	No data were available on wages and costs, but training seems to have been ineffective in terms of increasing the employability of participants.

Note: JSA, job search assistance, OJT, on-the-job training.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Addison, J. T., and W. S. Siebert. 1994. "Vocational Training and the European Community." *Oxford Economic Papers* 46(4):696–724.

Alfthan, Torkel, and B. Janzon. 1994. *Retraining Adult Workers in Sweden.* Training Policy Study 3 Geneva: International Labour Office.

Anderson, Patricia, Walter Corson, and Paul Decker. 1991. "The New Jersey Unemployment Insurance Reemployment Demonstration Project: Follow up Report." U.S. Department of Labor Unemployment Insurance Survey. U.S. Department of Labor, Washington, D.C., Processed.

Bloom, Howard. 1990. "Back to Work: Testing Reemployment Services for Displaced Workers." W. E. Upjohn Institute for Employment Research, Kalamazoo, Mich. Processed.

Corson, Walter, Stephen Long, and Rebecca Maynard. 1985. "An Impact Evaluation of the Buffalo Dislocated Worker Demonstration Program." Mathematica Policy Research, New Jersey. Processed.

Corson, Walter, S. Dunstan, Paul Decker, and A. Gordon 1989 "New Jersey Unemployment Insurance Reemployment Demonstration Project." Unemployment Insurance Occasional Paper 89-3. U.S. Department of Labor, Washington, D.C. Processed

Gill, Indermit S., and Amit Dar. 1995. "Costs and Effectiveness of Retraining in Hungary." Internal Discussion Paper IDP 155. World Bank, Europe and Central Asia Region, Washington, D.C. Processed.

Goss, Gilroy, and Associates. 1989. "Evaluation of the Job Development Program. Final Report Report prepared for the Program Evaluation Branch, Canadian Employment and Immigration Office, Ottawa. Processed.

Grossman, Jean B. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *The World Bank Research Observer* 9(2):159–80

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153–62.

———. 1992. "Randomization and Social Policy Evaluation." In Charles Manski and Irwin Garfinkel, eds, *Evaluating Welfare and Training Programs*. Cambridge, Mass., Harvard University Press.

Johanson, Richard. 1994. *Retraining Adults in Germany* Training Policy Study 4 Geneva International Labour Office.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76(4):604–20

Leigh, Duane. 1992. "Retraining Displaced Workers: What Can Developing Nations Learn from OECD Nations?" Policy Research Working Paper 946. World Bank, Population and Human Resources Department, Washington, D.C. Processed.

———. 1994. *Retraining Displaced Workers: The U.S. Experience* Training Policy Study 1. Geneva International Labour Office.

Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *The World Bank Research Observer* 9(2):181–201.

OECD (Organisation for Economic Co-operation and Development). 1991. *Evaluating Labor Market and Social Programs: The State of a Complex Art.* Paris.

———. 1993a. *Partnerships: The Key to Job Creation. Experiences from OECD Countries.* Paris.

———. 1993b. *Employment Outlook.* Paris.

———. 1994a. *The OECD Jobs Study: Evidence and Explanations.* Paris.

———. 1994b. *Vocational Training in the Netherlands: Reform and Innovation.* Paris.

O'Leary, Christopher. 1995. "An Impact Analysis of Labor Market Programs in Hungary." W. E. Upjohn Institute for Employment Research, Kalamazoo, Mich. Processed.

Pulay, Gyula. 1995. "Measuring the Efficiency of Employment Programs in Hungary." Paper prepared for a seminar on Evaluating Active Labor Programs. World Bank, Eastern Europe Department, Washington, D.C. Processed.

Revenga, Ana, Michelle Riboud, and Hong Tan. 1994. "The Impact of Mexico's Retraining Program on Employment and Wages." *The World Bank Economic Review* 8(2):247–77.

World Bank. 1995. *World Development Report 1995: Workers in an Integrating World.* New York: Oxford University Press.

Public Sector Reform in New Zealand and Its Relevance to Developing Countries

Malcolm Bale • Tony Dale

Does New Zealand's success story have lessons for developing countries contemplating public sector reform? That question usually elicits one of two reactions, both inadvisable in the authors' view. The first reaction is to be impressed with the efficacy of the reforms and conclude that they should be adopted uncritically in other countries. The second reaction is that the special conditions existing in New Zealand are such that none of its reform experience is relevant to others. The authors take a middle position, maintaining that poorer countries can indeed extrapolate from the experience of their higher income neighbor despite the different conditions under which they have to operate. New Zealand's comprehensive overhaul of its public sector affords both general principles and specific elements relevant to countries looking to improve the quality, efficiency, and cost effectiveness of their public service sectors, and a careful analysis of those reforms can ascertain what might be transferable and what principles might apply.

In 1984 New Zealand's newly elected Labor Government took over an economy characterized by comprehensive controls on the financial sector, extensive subsidies to farmers and exporters, and a highly sheltered private sector. Its deficit was a high 9 percent of gross domestic product (GDP), and public debt, at 60 percent of GDP, was rising. High underlying inflation and slow economic growth had reduced per capita income from one of the highest in the Organisation for Economic Co-operation and Development to one of the lowest (OECD 1983, 1984).

The new government put in place a macroeconomic stabilization plan and broad structural reforms to correct the core problems; the design, implementation, and outcome of the reform program are widely cited in the economic literature and business press and will not be elaborated here (OECD 1990–94; Bollard 1992; Evans, Grimes, and Wilkinson 1996). But government expenditures still accounted for about 40 percent of GDP, which, policymakers came to believe, meant that the improved

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 103–21

performance of the economy as a whole might be limited by the large public sector. Government departments were viewed as bloated, inefficient, and poorly managed. The bureaucracy behaved in ways that stereotypically undermine the capability of any government, and service delivery was poor. Departments habitually exceeded their budgets; unused balances were spent in end-of-year shopping sprees; creative accounting was used to give the appearance of good performance; and the management of assets and cash was haphazard.

Most important, politicians felt that some core ministries had their own policy agendas and could override or outlast the wishes of elected officials. Thus the objective of the new government was to create an efficient public sector that was also responsive to the strategic policy direction of the government.

The first step was to decide which activities the government should provide and which could be divested or spun off to the private sector. The second step was to undertake structural and management reforms in the remaining "core" departments—those concerned with broad, cross-cutting, nonsectoral issues such as Treasury, the ministries of Defense and Commerce, and the Inland Revenue Department (for details on the privatization program, see Boston and others 1991; Scott and Gorringe 1991). To paraphrase Holmes and Shand (1994), the government wanted the ethic of value for money and customer service to take its place alongside the ethic of probity and stewardship in the public sector.

Reform of Government Commercial Enterprises

In 1984 the government owned much of the economic infrastructure, including banking, postal, and telecommunications services; a steel mill; a shipping company production forests; electric power; and a large highway construction business. Most of these activities were being run by departments that also had policy advice func tions; nearly all ran at a loss and required taxpayer support. To reform the provision of these services, the government decided to "corporatize" these activities.

Corporatization involved forming government-owned enterprises with clear com mercial objectives, a neutral policy environment, managerial flexibility and author ity over decisions, performance monitoring, and explicit transfers for noncommer cial activities required by government to meet social objectives such as keeping ope small, unprofitable post offices or branch railway lines.

To meet these objectives, the enterprises were established under the normal com mercial company legislation common in all countries. The only difference from . regular commercial company was that the stock of the state enterprises was not pub licly traded but held by government ministers. These firms were required to operat profitably, to take on normal commercial levels of debt from the financial market and to pay taxes and dividends at commercial rates. Regulatory reform proceeded i

parallel with these reforms to ensure, wherever possible, competition in both the input and output markets. The companies were given the opportunity to succeed or fail based on their commercial performance.

Although corporatization was very successful, it had several drawbacks: the new corporations tied up capital that could be used to repay public debt; they were not subject to the ultimate discipline of the market because of the perception that an implicit government guarantee existed; the large amounts of capital needed to develop the businesses typically came at the expense of government investments in social infrastructure, such as health and education; and the government was exposed to commercial risk if the business suffered losses.

Thus successive governments have moved to sell their interest in commercial activities, a relatively straightforward procedure under commercial law. The government had a single criterion in each sale—to maximize the value to the taxpayers by selling the enterprise in a competitive environment. By mid-1997, 35 companies had been sold, at a value of approximately $15 billion. Most of the proceeds have been used to retire sovereign debt. The privatized entities have prospered, benefiting from the additional capital available from the private sector. Telecom New Zealand, for instance, went from one of the most inefficient telephone systems in the world to a position of international leadership in telecommunications services (Duncan and Bollard 1992).

All the remaining government-owned companies now run as profitable businesses and pay substantial dividends and taxes. For example, the quality of services provided by New Zealand Post has improved greatly. Although almost 80 percent of its revenue in 1994 came from activities that are competitive, it paid $22 million in taxes, paid dividends of $55 million to the government, and posted net profits of $46 million. In 1995 it lowered the cost of first-class postage from 29 cents to 25 cents and in 1996 offered users a "postage free" day. In 1997 its exclusive franchise over first class mail was removed so that its business is now fully contestable.

Reforming the "Core" Public Sector

To reform core activities, planners took their inspiration from public choice theory, principal-agent theory, transaction-cost theory, and the new public management literature (a brief explanation of these theories and a bibliography are provided in the appendix.) Agency theory was perhaps the most powerful of these concepts. This literature addresses the nature of contracts between two parties: the principal (or government) and the agent (or bureaucrats). In New Zealand planners believed that the problem was not with the bureaucracy—the civil service had many well-qualified and capable managers and staff who were responding in a rational way to the set of incentives they faced—but with the incentives themselves. The reforms were there-

fore intended to replicate, as closely as possible, the types of incentive structures for performance that might be found in a well-functioning private-sector concern, while taking into account the distinctive character of public services. The approach had five characteristics:

- Establishing clear lines of accountability between government ministers and their departments
- Defining performance in an unambiguous and measurable way
- Delegating authority to chief executives
- Establishing incentives that reward or punish results relative to the agreed outcome
- Reporting and monitoring performance.

Two laws were passed to effect these changes: the State Sector Act of 1988, and the Public Finance Act of 1989.

Accountability and Employee Relations

The emphasis on a strong accountability framework implied a new relationship between government ministers and permanent heads of departments. First, department heads lost their permanent tenure and were appointed to fixed terms up to five years, renewable for a further three years depending on performance. Now known as chief executives, department heads work under specified, performance-based contracts that they negotiate with the responsible minister. The State Services Commission monitors performance on behalf of the minister. The chief executives are free to run their departments in the ways they deem best suited to meet the performance goals. It is said that chief executives in the New Zealand public service used to have full tenure and limited authority; now they have limited tenure and full authority (Ball 1994).

Second, public sector employees were placed under the same rules and regulations as the private sector. That is, there are no public service employment regulations; all sectors, public or private, in New Zealand must comply with the same general employment regulations. The chief executive of each department is free to operate under more or less the same conditions as a chief executive officer of a private company making all decisions on the number of employees needed and the skill mix of personnel as well as all appointments and terminations (based on performance-based contracts for managers and staff).

Defining Performance

Because vague or unachievable performance specifications undermine good accountability arrangements, it was necessary to define the performance that departments chief executives were expected to deliver. Four elements of departmental performance were considered.

THE OUTPUT–OUTCOME DISTINCTION. In New Zealand accountability between ministers and their departments is based on the conventional distinction between outputs (goods and services produced) and outcomes (the effect of those outputs on the community). Chief executives are responsible for specified outputs from their departments, while the minister chooses which outputs will be purchased to achieve certain outcomes. That is, the minister, not the department, is responsible for the outcome. The distinction is important. Governments are interested in achieving outcomes and would like to contract for them if it makes sense to do so. Outcomes are often not within the control of the chief executive, however, and he or she cannot be held accountable for them. But chief executives can be held accountable for outputs, which can be relatively well-defined and are within the executive's control.

Say, for example, the police commissioner contracts with the minister of police to provide a certain level of policing services, patrols, community security programs, road safety commercials, and so forth. These are clearly outputs. The commissioner does not contract to lower the crime rate. That outcome may be forthcoming, but the crime rate is affected by many variables beyond the control of the commissioner, such as the level of unemployment, immigration policy, social policy, perhaps even the result of certain sporting events. Thus, holding the commissioner accountable for outcomes is not operationally useful; this distinction between who is accountable for outputs and who for outcomes is common in the private sector as well.

Accountability processes are much more effective if the outputs to be delivered are well specified in advance. Vague specifications allow managers to determine exactly what it is their organization will produce. For this reason, the minister and chief executive prepare annual purchase agreements, or contracts, that set out in reasonable detail the outputs to be delivered. The quality and robustness of this system depend on the careful specification of outputs in the agreement. Many ministers have purchase advisers to assist them with this task.

A common concern is that this approach emphasizes outputs at the expense of outcomes. Critics maintain that unless departments are held accountable for outcomes, they will not focus on these issues and therefore are less likely to achieve them. That is not necessarily the case. Greater clarity over what is being produced (outputs) can increase the attention to outcomes. Since 1990 the government, as part of the annual budget process, has reviewed all the outputs produced by departments against the criterion: How does this help achieve the outcomes the government wishes to pursue? This focus makes clear that policy advice should be about the relationship between interventions (including outputs) and outcomes, about what the government is trying to achieve and alternative ways to achieve it. This framework has helped departments understand that, just as in the private sector, their survival is dependent upon meeting the needs of their customer. Because their customer is interested in outcomes, departments, given sufficient competitive pressure, will strive to design and provide public services to help achieve those outcomes.

THE OWNER–PURCHASER DISTINCTION. In addition to purchasing most of a department's outputs, the government is also the department's owner. As the purchaser, the government wishes to obtain goods and services of a specified quantity and quality for the lowest price, whether it is buying from the private or public sector. To enhance price competition, all ministers are free to purchase their outputs from nondepartmental sources. The finance minister, for example, may purchase economic projections and advice from Treasury officials or from any private domestic or international economic consulting firm. As the owner of the Treasury, the government wants to obtain the best possible return on its investments in it. If the Treasury fails to record a positive rate of return, public production lowers the wealth of society. It would better in that case to close that department and contract out all of its work (outputs).

The New Zealand approach recognizes these two perspectives as two differen dimensions of departmental performance. Performance agreements with each chie executive separately specify each dimension, and both are monitored. The fact tha governments purchase goods and services from their departments and also own thos departments could give rise to conflicting objectives. This conflict is solved by pric ing departmental outputs at a price equivalent to that charged by the private secto The department's performance as a business can then be fairly assessed using norm business evaluations that enable governments to determine whether the departme is operating efficiently and whether owning the business is advantageous.

Where the private sector produces the same output, this model does not pose problem. But where there is no comparable private supplier, an alternative benc mark must be established. In some cases, prices charged by different public sect entities can be compared, as in a 1993 study of the unit price of policy advice ou puts provided by departments. Where no endogenous benchmark can be establishe officials are experimenting with establishing an initial price that applies some prici rule to costs (such as cost minus *x*). Although this method will not establish efficient market price, it does create some separation between cost and price.

Ministries regard the minister as a customer rather than as the recipient of a s vice. Ministries are accountable *only* to their minister, not to taxpayers and servi recipients. Ministers are accountable to taxpayers and service recipients. This co trasts with the citizen charter approach used in the United Kingdom, which crea a dual accountability for departments: to their minister and to the recipients of th services. A difficulty can arise if those accountabilities are inconsistent —for examp if the quantity or timeliness of services demanded by recipients is incompatible w the amount ministers are prepared to pay for the services. The approach used in N Zealand resolves this problem by having ministers make the tradeoff among qu tity, quality, time, place, and cost of service delivery; that is, between common int est and individual interest. This does not mean that departments are not concer about whether customers are satisfied. Such a response will inevitably becom

concern of the minister, whose political antennae should be well attuned to such feedback.

THE GOVERNMENT–DEPARTMENT DISTINCTION. Chief executives make all input decisions, including capital investment decisions (within a defined capital base). Indeed, such authority is necessary if they are to be held accountable for producing outputs in the most efficient manner. Some departments, however, manage inputs over which they do not have full control, that are not used to produce their outputs, or for which the government does not wish to delegate authority. As a result, activities and assets have been divided into two categories: "the Crown" (ministers); and the departments. For Crown activities or assets, accountability remains with the government. For example, the Department of Social Welfare is responsible for administering the social support scheme, but the minister is responsible for the size and number of welfare payments. Understandably, the chief executive cannot be held accountable for, say, the number of people unemployed or the number of single-parent children. Similarly, the national parks are managed by the Department of Conservation for the government but are owned by the government and listed as a *Crown asset.*

POLICY ADVICE–SERVICE DELIVERY DISTINCTION. Where an agency provides both policy advice and service delivery, a potential conflict of interest arises between the two functions. Separating them reduces the potential for policy advice bias. Under the reforms, policy advice is an output provided by departments in much the same way as consulting advice is an output provided by a consulting firm. Good policy advisers must be able to evaluate the tradeoffs between different outcomes and identify the nexus between outputs and outcomes. They must evaluate spending proposals against all alternative interventions that could produce the same outcome. If, for example, the government decides it wants to reduce road accidents, it may purchase outputs such as highway patrols, road repairs, and vehicle inspection checks. Alternatively, it may intervene legislatively with speed limits, higher alcohol taxes, different speed sanctions, or compulsory driver-education programs.

Viewed from that perspective, policy advice is a specialized business that is inherently different from service delivery. In most cases, it has been decided that policy advice is most effectively and efficiently produced by a department dedicated to its production rather than by one with distinctly different lines of business. For example, in the environmental area, policy advice is the responsibility of the Ministry for the Environment, while another agency, the Department of Conservation, is responsible for delivering services such as operating the national parks. Because the ministry does not deliver services, its advice about appropriate interventions can be independent of the business implications for the department.

A related reason for separating these functions is to reduce the tendency for specia interest groups to "capture" the agency that regulates them (Posner 1974). Thi rent-seeking behavior can be reduced if regulatory policy is designed in one agenc and enforced in another.

Reporting, Monitoring, and Coordination

Improved reporting and monitoring of departmental performance was the *quid p quo* for enhancing the chief executive's autonomy. This required upgrading financi management systems and skills.

EX POST REPORTING. Defining and monitoring purchase and ownership performan requires information about the full resource cost, including the consumption of a sets and the opportunity cost of capital, and about assets and liabilities, their utiliz tion, and the return being generated. For this reason, all government entities a required to report financial performance on an accrual accounting basis, using tl same generally accepted accounting practices as does the private sector.[1]

Each department must provide a full set of financial statements to its minister ar to the Treasury on a monthly basis. In addition, departments must produce ar submit an audited annual Statement of Service Performance, outlining the outpu produced versus the outputs agreed and giving information about purchase perfo mance in the same way that private companies produce annual reports showing th financial statements and performance. As a result the government can simply sum of the accounts to produce national financial statements on an accrual basis. The are published monthly, and annual audited financial statements are presented to t Parliament within three months of the close of the fiscal year. Thus governme accounts look similar to an annual report produced by a private company. Th show, among other things, the net worth of the government.

EX ANTE BUDGETING. The reforms also changed the budget and appropriations s tems in two ways to fit with the performance management system. First, appropr tions for departmental outputs (not inputs) are now made on an accrual accoun basis, and managers are free to acquire their inputs from any provider. Second, o capital injections into departments are appropriated (not capital expenditures), flecting the chief executive's authority to manage assets within a defined capital ba Budgetary reporting at both departmental and national levels mirrors ex post repo ing. Financial forecasts are prepared using generally accepted accounting practi and are identical in form to financial statements produced by private corporatio

CAPITAL CHARGE. Twice a year departments pay a capital charge, calculated on basis of their net assets, for the cost of the capital the government has investec

them. This charge has several benefits. First, it ensures that the cost of capital is reflected in output prices because a department's total cost must be allocated to its outputs to ensure comparability with nongovernment producers. Second, the charge encourages departments to manage their balance sheets carefully and to divest surplus or redundant assets. Third, it encourages management to consider the mix of assets needed to produce services efficiently. If a manager finds it more efficient to purchase more computers and sell some cars, he or she is free to make this decision. If the sale of assets reduces the overall capital charge, the savings can be applied to other expenses.

STRATEGIC POLICY COORDINATION. Three means were devised to coordinate strategic policy, which was a key objective of the reforms. Ministers were required to specify and publish the outcomes the government wished to achieve. Coordinating committees were established, made up of senior government officials from departments with an interest in broad areas of government activity, such as social policy, education and training, or environment. Their role is to ensure that policy options are developed in a coordinated way across the government. The third method of coordination is created by the policy-delivery split. Because ministers want advice about policies that can be effectively implemented, and they want output proposals that meet their objectives, policy and delivery departments have a relatively strong incentive to coordinate with each other. Rather than weaken the horizontal linkages between departments, these arrangements have served to strengthen them.

The use of the private sector profit center approach combined with market pricing techniques has also solved one of the common public sector problems: transfer pricing—the price of a good or service for a transaction within an organization. Because there is no market price, the organization can set the price at any level, including zero. Modern business practices require that transfer prices approximate market prices so that accurate profit-loss data can be established by each unit of an organization. Because each department is regarded as a profit center and all transactions between departments are treated as arms-length transactions, the transfer pricing problem has effectively disappeared.

Results

The core sector reforms have succeeded in improving both service delivery and efficiency. The system is widely supported by departmental managers, although the effect of the reforms on managerial behavior has varied depending on the quality of leadership and the levels of efficiency prior to the reforms. In general, performance has improved in tandem with the development of wage scales linked to performance. Savings from improved cash management have been substantial (enough to pay for all the system costs of the reforms), and unappropriated expenditures, which were

Box 1. Reforming Health Care: Anatomy of Limited Success

Historically, New Zealand's health care system has been predominantly publicly operated an funded through general taxes. Before 1993, when the health care system was reformed, secondar and tertiary (hospital) services were the responsibility of area health boards, with funding provide according to population. The new system (which is similar to the United Kingdom's) is based on Ministry of Health (a policy advice agency), four regional health authorities (who purchase primar secondary, and tertiary health services for specified regional populations), and a series of Crow health enterprises (government-owned hospitals run on commercial business practices). Primai services continue to be provided by private practitioners (who are heavily subsidized). The region health authorities determine the mix of health services to be provided to their populations within given funding level. They are then responsible for purchasing those services and monitoring deliver with oversight from boards of directors made up of health, community, and commerci representatives.

The results have been mixed, but generally the public health system has failed to achieve tl performance gains expected, and considerable political and public debate continues about tl desirability of the arrangements. Although the quality of financial management and the degree transparency among the Crown health enterprises has improved significantly, efficiency gains ha been slow to emerge, and government funding has been required to keep many enterprises solve The public remains uneasy with the notion of for-profit hospitals. The explicit division of poli and purchasing responsibilities has been less successful than anticipated, health officials mainta significant control over purchasing decisions, and there is a continuing demand for additional fundir

As a result of political compromises, most of the reforms to improve the delivery of social servic such as health and education have not been implemented as designed. The reforms were predicat on the existence of market competition (on both the supply and demand side), which has n developed. The quasi-commercial and social objectives of providers have diluted the focus performance. Political issues were also underestimated. Social policy concerns the public rationi of private goods, and removing these concerns from the political arena has proved to be much mo difficult than anticipated. This has resulted in much more direct ministerial involvement (so would say interference), which has undermined the authority of the various providers. The contr in performance between the purer commercial model versus the political-interference model provi a salient lesson. If quantum efficiency gains are desired and if commercial objectives are to be achiev it is necessary to divorce political influence from commercial decisionmaking

and poor staff performance to be the problem. They were, in fact, only the sym toms; the real problem was the lack of management incentives. Second, the fran work provided consistency for the multiple layers of decisions required in the c sign and implementation of the reform. These decisions, such as the nature of t accounting system, the approach to budgeting, and the degree of personnel d egations, were all made in a consistent manner. Third, it focused attention oi comprehensive approach. As a result the reform addressed all aspects of pub sector management and all parts of the public sector (departments, governm corporations, local governments). Fourth, the framework guided the sequenci and implementation of the reforms. These decisions were based on what was m important from a top-down perspective rather than what took the fancy of depa

ments. Fifth, it aided in marketing the reforms to departments and the public. The coherence and comprehensiveness reduced fears that the reform was just another ad hoc initiative. The lesson for other countries is clear. Basing reforms on an analytically rigorous conceptual framework appropriate for the jurisdiction concerned and having the framework apply to the entire public sector is likely to improve significantly the chances that the reform will be successful.

The Importance of a Clear Performance Definition

The second lesson relates to performance specification. The reforms were based on identifying the various principal-agent relationships; specifying and reporting performance in a clear and unambiguous manner; and ensuring that managerial authority matched the department's responsibilities. Providing state-owned enterprises with a clear profit-maximizing objective is an example of this expectation. So, too, are the features described in the discussion of performance, which are designed to clarify what performance the principal expects of the agent in a way that holds the agent unambiguously accountable. However performance is defined, the agent must have control over it; for this reason, the outputs approach has great merit.

The existence of several principals (as in the U.S. separation of powers) does not necessarily invalidate the principal-agent approach. Rather, it highlights the need to be clear about the different aspects of performance each principal requires of the agency, and for the principals to reconcile any mutually exclusive conflicts in those requirements.

New Zealand sought to ensure that managers faced incentives congruent with the performance expected of them. This basic notion is often overlooked in the design of public sector reforms. Performance incentives are much more subtle and pervasive than the "personal bonus or reward" incentive commonly mentioned. Other aspects need to be considered, including

- Whether the institutional and organizational arrangements encourage the performance required.
- Whether agencies can be made more efficient and responsive to their customers' needs, either by increasing competition or by direct customer purchasing of services (rather than the minister purchasing services on behalf of consumers). The evidence is that they can.
- Whether all aspects of the management system are sending the same signals to managers. Often budget and accounting systems define and measure performance in totally different ways.
- Whether the actual performance of all the actors in the systems is transparent.
- Whether the system is empowering or controlling.
- Whether the systems encourage managers to actively manage all their resources. The purchase of cleaning services, motor vehicles, or even personnel by central

agencies at no cost to operating departments discourages those departments fro[m] efficiently managing their costs.
- Whether the personnel system encourages good performance.

Focusing on What Government Does Best

It is important to focus the government's resources on areas in which governme[nt] can add value, such as establishing an appropriate regulatory framework and ec[o]nomic environment in which the private sector can thrive. Wherever possible, Ne[w] Zealand turned over businesses to the private sector because governments do n[ot] have the ability to manage and monitor enterprises in the same thorough way as do the private sector. Commercial objectives and decisions are easily compromised [by] political and social ones. In addition, state enterprises compete for scarce gover[n]ment financial and management resources; thus they tend to be undercapitaliz[ed] and undermanaged. Governments should concentrate on what they do best—esta[b]lishing the regulatory and economic environment, financing public infrastructu[re,] and ensuring the delivery of public goods. And these activities can be operated alo[ng] commercial business lines to a much greater extent than is commonly believed.

A public sector management system is a means; not an end. It will not deli[ver] better fiscal performance on its own, but instead needs to work inside an overarch[ing] political economy framework that sets clear macroeconomic goals and has the pol[iti]cal resolve to achieve them. This point is clear from the New Zealand experience. [All] of the reforms were driven by the same central goal: to improve macroecono[mic] performance by reducing the negative impact of the public sector. A good pu[blic] sector management system will provide politicians with the tools they need to achi[eve] those targets; a poor one will make a politician's task more difficult.

It is more difficult to draw conclusions about whether specific practices ado[pted] in New Zealand are applicable to developing countries. If the cultural and polit[ical] environment is too dissimilar, the applicability of these practices may be limi[ted.] The following precedents formed the basis for the reforms adopted in New Zeal[and:] a tradition of a politically neutral, relatively competent civil service; little con[cern] about corruption or nepotism; a consistent and well-enforced legal code, inclu[ding] contract law; a well-functioning political market; and a competent, but suppres[sed] private sector.

The right reform mix for any developing country must reflect any major di[ffer]ences in these preconditions; New Zealand's reforms cannot simply be transpla[nted.] For example, because reducing corruption or nepotism and increasing demo[cratic] participation were not objectives in New Zealand, the reforms did not address t[hese] issues. If financial performance and service delivery are concerns, however, then [spe]cific New Zealand techniques may be appropriate. Subject to this caveat, the fol[low]ing techniques are likely to find broad applicability in developing countries.

- If the performance of government commercial activities is an issue, developing countries would find it useful to separate trading activities from core departmental functions, apply a "level playing field" regulatory regime, and appoint independent boards with business—rather than political—expertise. The success of such reforms is dependent on acquiring the necessary governance and managerial skills from the private sector. Many developing countries have this expertise.
- If the concern is about users of the service heavily influencing policy advice, then separating policy advice and service delivery functions into different agencies is likely to lead to both better advice and better service delivery because conflicting objectives have been removed. If this route is pursued, the institutional design should put more emphasis on processes to coordinate advice. These can operate at the political level, the bureaucratic level, or both.
- If service delivery is a concern within core government agencies, a management system focused on outputs would be relevant to developing countries. Politicians and managers would need to make a significant paradigm shift in the way the government operates, not least in securing the political commitment to such a change. Publishing the agreed outputs would be a first step in transparency and has the potential to improve accountability. To be fully effective, the planning, budgeting, appropriation, performance assessment, and reporting systems all need to be based on outputs. This does not preclude accountability for input management as well, if that is considered desirable.
- If financial performance is an issue, improved measurement of financial performance (through the use of accrual accounting) and explicit attention to this information in budgeting and reporting systems is important, although skill shortages may make implementation difficult for some countries. Significant gains can be achieved through improved financial management and financial control systems even without accrual accounting. The establishment of an effective cash accounting system may be considered a precursor to accrual accounting.
- If personnel performance is a concern, performance based personnel systems are likely to improve the staff's incentives to perform. Such contracts may be politically difficult, but they have been a central part of institutional reform in New Zealand, are viewed as very effective, and were the key to changing the public service culture. Even if this model is not fully acceptable in developing countries, changing the appointment, appraisal, and promotion systems so that they are based on performance is likely to be possible and beneficial.
- Transparency is an important incentive device. Providing that voters are able to replace their politicians if they do not make decisions in the public interest, developing countries can require improved specification and reporting of performance to government and to the public.

and recurrent, complex transactions. According to Williamson (1985), these factors explain the concentration of production in some sectors in a few large firms. Thus direct provision may be preferable when maintaining quality is critical and opportunism poses a serious threat. It is for these reasons that governments are hesitant to contract out the gathering of military intelligence and the collection of taxes.

The New Public Management

This approach (see Aucion 1990 and Caiden 1988) centers on the presumption that a distinct activity called "management" can be applied to public and private businesses alike, and that it includes the following elements: a move away from input controls, rules, and procedures toward output measurement and performance targets—the "accountability" framework; the devolution of management control with improved reporting and monitoring mechanisms; a preference for private ownership, contestable provision, and contracting-out of publicly funded services; the adoption of private-sector management practices in the public sector, such as short-term labor contracts, performance-linked remuneration schemes, the development of a mission statement, greater concern with corporate image, and the development of a corporate strategy and action plan; an emphasis on efficiency, often referred to as "value for money" (Hood and Jackson 1991).

Notes

Malcolm Bale is principal economist, East Asia and Pacific Region, the World Bank, and Tony Dale is a principal of Public Sector Performance (NZ) Ltd., a public management consulting business

1. Accrual accounting is an accounting method that recognizes transactions and other events when they occur and not as cash transactions or their equivalent. The events are recorded in the accounting period and reported in the financial statements in the periods to which they relate

2. The five principles of responsible fiscal management are (1) reducing total Crown debt to prudent levels by ensuring that total operating expenses for the Crown are less than total operating revenues in the same financial year; (2) maintaining prudent debt levels, once they have been achieved by ensuring that on average over a reasonable period of time, Crown operating expenses do not exceed Crown operating revenues; (3) achieving and maintaining Crown net worth at a level that provides a buffer against future adverse events; (4) managing prudently the fiscal risks facing the Crown; and (5) pursuing policies that are consistent with reasonable predictability about the level and stability of tax rates.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Aucoin, Peter. 1990. "Administrative Reform in Public Management: Paradigms, Principles, Paradoxes, and Pendulums." *Governance* 3:115-37.

Ball, Ian. 1994. "Reinventing Government: Lessons Learned from the New Zealand Treasury." *The Government Accountants Journal* (Fall).19–28.

Bendor, Jonathan. 1988. "Formal Models of Bureaucracy." *British Journal of Political Science* 18(3):353–95.

Bollard, Alan. 1992. *New Zealand Reforms, 1984–91.* San Francisco: ICS Press.

Boston, Jonathan, John Martin, June Pallot, and Pat Walsh, eds. 1991. *Reshaping the State: New Zealand's Bureaucratic Revolution.* Auckland: Oxford University Press.

Buchanan, James, and Gordon Tullock. 1962. *The Calculus of Consent. Logical Foundations of Constitutional Democracy.* Ann Arbor: University of Michigan Press.

Caiden, G. 1988. "The Vitality of Administrative Reform." *International Review of Administrative Sciences* 54(3):331–58.

Deloitte, Ross, Tohmatsu. 1990. "Post-Implementation Review of Cash Management Reform." Paper prepared for and available from the New Zealand Treasury, Auckland. Processed.

Duncan, Ian, and Alan Bollard. 1992. *Corporatization and Privatization: Lessons from New Zealand.* Auckland: Oxford University Press.

Evans, Lewis, Arthur Grimes, and Bryce Wilkinson. 1996. "Economic Reform in New Zealand, 1984–95: The Pursuit of Efficiency." *Journal of Economic Literature* 34(December):1791–902.

Holmes, Malcolm, and David Shand. 1994. "Management Reform: Some Practitioner Perspectives on the Past Ten Years." In *Ten Years of Change: Proceedings of the SOG Ten-Year Reunion.*

Hood, Christopher, and Michael Jackson. 1991. *Administrative Argument.* Aldershot, U.K.: Dartmouth Publishing.

Moe, Terry M. 1984. "The New Economics of Organization." *American Journal of Political Science* 28:739–75.

———. 1990. "Political Institutions: The Neglected Side of the Story." *Journal of Law, Economics, and Organizations* 6:213–53.

Niskanen, William. 1971. *Bureaucracy and Representative Government.* Chicago: Rand McNally.

OECD (Organisation for Economic Co-operation and Development). 1983, 1984, 1990–94. *OECD Economic Surveys: New Zealand.* Paris.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge, Mass.: Harvard University Press.

Posner, Richard A. 1974. "Theories of Economic Regulation." *Bell Journal of Economics and Management Science* 5(2):335–58.

Pratt, John, and Richard Zeckhauser, eds. 1985. *Principals and Agents: The Structure of Business.* Boston: Harvard Business School Press.

Scott, Graham, and Peter Gorringe. 1989. "Reform of the Core Public Sector: The New Zealand Experience." *Australian Journal of Public Administration* 48(1):81–92.

Tullock, Gordon. 1965. *The Politics of Bureaucracy.* Washington, D.C.: Public Affairs Press.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting.* New York: Free Press.

Why Most Developing Countries Should Not Try New Zealand's Reforms

Allen Schick

During the past decade New Zealand has introduced far-reaching reforms in the structure and operation of government departments and agencies. This model has attracted interest in developing countries because it promises significant gains in operational efficiency. But developing countries, which are dominated by informal markets, are risky candidates for applying the New Zealand model. The author suggests that basic reforms to strengthen rule-based government and pave the way for robust markets should be undertaken first.

Developing and transitional countries have an understandable desire to accelerate public sector reform by adopting the most advanced innovations devised by industrial countries. This interest has been stimulated by the New Zealand model, which gives public managers broad discretion to operate within an accountability framework that specifies the results to be achieved and closely monitors performance. During the past decade, dozens of countries have sent delegations to New Zealand to observe its avant garde management practices and to interview government officials on how the new systems and procedures have affected the cost and delivery of public services. The World Bank and other international organizations have showcased New Zealand's reforms at various conferences, and some of the architects of the reforms have crisscrossed the globe extolling the virtues and portability of their country's version of results-oriented public management.

Despite the interest and the sales efforts, only a few developed countries (such as Iceland and Singapore) have adopted selected features of the model; others (such as Sweden and the United Kingdom) have embraced a managerial ethic without subscribing to the hard-edged contractualism that differentiates New Zealand's reforms from those tried elsewhere. To this writer's knowledge, however, not a single developing or transitional country has installed the full New Zealand model, although quite a few have been enchanted by the prospect of leapfrogging to the front ranks in the international reform sweepstakes. A few countries (such as Mongolia) are in the

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 123–31

early stage of adopting selected features of the system, but it is much too early to gauge how far they will go in embracing its basic tenets. On the whole, industrial and developing countries have not implemented such reforms because the reforms are beyond their reach or do not fit their current needs.

I draw this conclusion despite New Zealand's enormous contribution to the theory and practice of public management. Not only has the menu of reform possibilities been greatly expanded, but New Zealand has brought its public management much more closely into line with institutional economics and with contemporary business practices. And the rigor with which the model has been applied is impressive. This was not a case in which reformers selected discrete entries from a large menu of reforms. The change agenda was driven by ideas that have only recently entered mainstream economics, and the ideas were applied with full fidelity to their internal logic. But I do not accept the view that New Zealand offers practical guidance on how developing countries should surmount deficiencies in public management.

I believe that there are important preconditions for successfully implementing the new public management approach and that these should not be ignored by countries striving to correct decades of mismanagement. In contrast to those who take the position that managerial deficiencies should be the driving factor in determining the suitability of these type reforms, I argue that they should be deterring factors. The greater the shortcomings in a country's established management practices, the less suitable the reforms.

The New Zealand Model: Government by Contract

Since 1988, New Zealand has implemented an enormous number of management reforms that add up to an integrated concept of how government should work. That concept is expressed in the heading to this section. Virtually every element of reform has been designed to establish or strengthen contract-like relationships between the government and ministers as purchasers of goods and services and departments and other entities as suppliers. Hundreds of contracts are formally negotiated each year; the typical contract specifies the resources that one side will provide and the performance the other side will produce. Ministers are always on the resource-providing side of the relationship; chief executives can be on either side, depending on the role they are playing. A chief executive provides resources in negotiating employment contracts with managers but promises results in negotiating purchase agreements with ministers and performance agreements with the State Services Commissioner.

This "new contractualism" replaces the implicit or relational contracts that characterize traditional public administration.[1] Contracts convert the budget from an understanding between government and parliament on the amounts to be raised and spent into an explicit statement of what will be done with the resources to be made

available. In a similar vein, performance agreements displace the old civil service ethic of trust and responsibility with accountability for the results expected from each chief executive.

New Zealand has gone to extraordinary lengths to create conditions under which formal contracts are negotiated and enforced. It has restructured many departments to decouple policymaking functions from the delivery of services. (For example, the national defense organization was split into two separate entities: the Ministry of Defense, which provides policy advice to government; and the Defense Forces, which carry out assigned operations). Under the new system, ministers can purchase services from government departments or from any alternative public or private supplier. Appropriations are on an accrual basis, so that the full cost of the goods and services is incorporated in the purchase price. In fact, to promote competition among suppliers, appropriations include an amount for the tax on goods and services that would be paid in a private transaction. In a similar vein, a capital charge is levied on the net worth (as shown on its balance sheet) of each department. This charge reflects the opportunity cost of money and resembles the internal rate of return expected by firms from their operating units.

Contract-like arrangements have been extended to policy advice as well, so that ministers can opt to obtain information and ideas from consultancies and other external sources. To put alternative suppliers on an equal footing, the government accords the chief executives of departments the same operating flexibility that is enjoyed by executives in nongovernmental organizations. Chief executives are given a block of resources for each class of outputs they contract to purchase, and they have discretion to select the mix of inputs used in producing the outputs. These outputs are specified in detail so that the government can have reasonable assurance that departments are producing the outputs contracted for. Under the present system, when the budget is tabled in Parliament, each department publishes a departmental forecast report that, among other things, specifies the outputs it will produce in the next financial year. Shortly before the start of the year, the outputs are specified in greater detail in purchase agreements signed by the chief executive and the minister purchasing the services. Multiple purchase agreements are written when more than one minister purchases services from the same department. After the year is done, each department publishes an annual report that specifies the outputs actually produced, thereby enabling the government to determine whether the terms of various contracts have been fulfilled.

Before assessing whether the New Zealand model is appropriate for developing countries, it is necessary to consider its effectiveness at home. In my view, organizational performance has been significantly enhanced. But this favorable assessment carries certain caveats, some of which were discussed in my report commissioned by the New Zealand Government (Schick 1996).

- The New Zealand model emphasizes matters that can be specified in contracts, such as the purchase of outputs, but gives inadequate attention to outcomes and the government's ownership interest because they do not fit easily into the contracting framework.
- Robust contracting depends on voluntary, self-interested action. Sometimes, however, self-interest defeats the government's collective interest. In the early years of reform, for example, efforts to establish a senior executive service were undermined by managers who preferred to contract on an individual basis.
- Contractualism may weaken traditional values of public service, personal responsibility, and professionalism. It can induce managers to take a checklist approach to accountability—"if it's not specified, it's not my responsibility."
- Contract-like arrangements do not themselves create arms-length relationships in the public sector, nor do they enable the government to toughen its insistence on performance. In most cases, government has little choice but to contract with internal suppliers, typically its own departments. If these fail to perform, the government can sack the chief executive and apply some pressure. But it rarely has the exit option that is essential to the effectiveness and enforcement of private contracts.
- Chief executives, senior managers, and others attribute most of the improvement in government performance to the discretion given to managers rather than to formal contracts. Managers differ on how much value is added by contracts, but few think that they have been the main contributor to higher operational efficiency.
- Contracting is not costless. Negotiating and enforcing contracts entails enormous transaction costs that have not been systematically studied, although they take a deep bite out of operating budgets, especially those of small departments

These concerns point to the unfinished business of public sector reform in New Zealand. There is much more to be accomplished before a final assessment can be made. At this early stage, one is justified in acknowledging that the country has vastly enlarged the stockpile of public management ideas and practices. In promoting internal markets within government, it has devised creative alternatives to privatization while carrying the pursuit of operational efficiency well beyond standard market-type mechanisms such as user charges.

Yet one should not lose sight of the fact that these are not real markets and that they do not operate with real contracts. Rather, the contracts are between public entities: the owner and the owned. The government has weak redress when its own organizations fail to perform, and it may be subject to as much capture in negotiating and enforcing its contracts as it was under pre-reform management. My own sense is that while some gain may come from mimicking markets, anything less than the real thing denies government the full benefits of vigorous competition and economic redress.

The Informal Public Sector

In New Zealand, formal contracts and internal markets were feasible because the country had a robust market sector and established mechanisms for enforcing contracts—conditions that are often absent in developing countries, which tend to have an informal economy with relatively weak specification of property rights and other formal processes to regulate economic activity.

Informality is not a new concept; shortly before New Zealand embarked on its reforms, de Soto (1989, p. 12) emphasized informality as the distinctive condition of the Peruvian economy. Characterizing it "as a gray area which has a long frontier with the legal world and in which individuals take refuge when the cost of obeying the law outweighs the benefits," de Soto found that the informal economy supplied more housing to Peruvians than did the government, was the main source of public transportation, and enabled entrepreneurs to start businesses when they were blocked by government regulations. He also concluded that the informal economy was inefficient, bred corruption, denied home and business owners access to capital, and retarded economic development.

De Soto and others have focused on informality in the market economy. I believe that informality is as pronounced in the culture of government as it is in the marketplace. In fact, the parallel incidence of informality in the public and private sectors is not happenstance. Norms, practices, and ideas migrate from one sector to the other, as does the dead hand of overregulation and the eagerness of government officials to look the other way in exchange for favors. The emergence of open, robust markets is as much a precondition for modernizing the public sector as it is for developing the private economy. It is highly unlikely that government will operate by the book when rules and regulations are routinely breached in private transactions. If New Zealand style contracts are at one end of the spectrum, then informality is at the other end. And if contracts and the rule of law are underdeveloped in business relations, it is highly improbable that they can be effectively applied in the conduct of the government's business. It would be foolhardy to entrust public managers with complete freedom over resources when they have not yet internalized the habit of spending public money according to prescribed rules. Many developing countries have formal management control systems that prescribe how government should operate. These systems are overseen by powerful central agencies such as the finance ministry, the civil service board, and the procurement agency. On paper everything is done according to rule. The civil service system is based on a detailed classification of positions and ranks, each with its own job descriptions, skill and experience requirements, eligibility rules, and pay scale. In this formal control process, operating units must obtain advance approval from the civil service agency (and sometimes from the finance ministry as well) before they can fill vacant positions. Formal rules dictate every step in the hiring process: announcing the position, establishing eligi-

bility qualifications, processing applications, appointing the winning candidate, and setting each employee's pay and grade levels. Each step is monitored by a central agency to assure compliance with the rules.

Where informality flourishes, however, this is not the way many civil servants get their jobs. They are hired because they know the right person or have contributed to some organization or cause. Because official pay levels are low, they may be assigned to one position but be paid for another. Many may be ghost workers who appear on the payroll but not at work; some may hold two or more positions, and those who show up on the job may put in less than a day's work because the official salary scale is a lot less than a day's reasonable pay. Thus there are two coexisting civil service systems—one based on formal rules, the other on actual practices. To say that there is an informal system is not to conclude that the rules always are ignored or that corruption always flourishes, although these pathologies may occur. Rather, it is to argue that the informality contributes to public order; in the case of the civil service it enables the government to recruit and retain skilled persons.

Informality also reigns in the budget arena. The government has two budgets: the public one that is presented to the parliament and the real one that determines which bills are paid and how much is actually spent. The formal budget promises spending that exceeds the government's fiscal capacity; the informal budget facilitates macroeconomic stability by not making some of the expenditures approved by the parliament. The formal budget is known in advance, the informal one after the spending occurs. Because there are two budgets, the temptation is for the formal document to be unrealistic and unachieveable. The process thus feeds on itself. Inasmuch as the official budget will not be implemented, why not cram into it spending that will not be made? This behavior leads to cash flow budgeting, in which the amounts actually spent are determined more by cash payments than by the amounts authorized by law. It also breeds repetitive budgeting, in which the government "rebudgets" several times during the year to align disbursements and resources.

Informality is a mixed blessing. On the one hand, it cuts through red tape, unresponsive bureaucracies, and bad policies; on the other hand, it opens the door to (and sometimes institutionalizes) corruption and inefficiency. The positive side of informality in public management includes the maintenance of fiscal discipline despite unrealistic budgets and the provision of public services despite rigid rules and controls. But the costs are high; they include widespread evasion of civil service rules and other controls, the time and resources spent in beating the system, distrust of government, routinized corruption, and inattention to the outputs and results of public programs and the performance of government agencies and officials. It would not be surprising if some of the most esteemed and productive civil servants in developing countries are those who use their entrepreneurial and managerial skills to outwit the formal controls. But when bureaucrats are valued for their verve in operating informally, it is easy for them and others trapped in the system to lose sight of the

public purposes they are serving and the outputs they are supposed to produce. It is only a short step from disabling the controls to bending the rules for dishonorable purposes.

It is a much longer step for them to adopt New Zealand–style reforms, and a much riskier one. No country should move directly from an informal public sector to one in which managers are accorded enormous discretion to hire and spend as they see fit. New Zealand did not make this leap, and neither should other countries. Before reform New Zealand operated under budgets that controlled spending and corresponded to actual transactions; it also had a civil service system that governed how public employees were hired and paid. In other words, it had a formal public sector. This is an essential precondition for adopting elements of the New Zealand model.

The Logic of Development

If contract-based public management is beyond reach and informality is an unsatisfactory state of affairs, what can developing countries do to improve government operations? In my view, significant progress can be made through a logical sequence of steps that diminish the scope of informality while building managerial capacity, confidence, and experience. This concluding section outlines some of the key steps.

First, progress in the public sector requires parallel advances in the market sector. As long as the economy operates according to informal norms and property rights are defined more by practice than by contract, the government is not likely to make much headway in installing rule-based public management. There may be special situations (in colonial regimes, for example) that enable a developing country to establish a skilled civil service system, modern financial management, and other trappings of formal public management even though the market sector is lagging behind. But the much more typical situation is one in which market development precedes or coincides with the development of robust public institutions. Singapore and Chile are countries in which economic development and modernization of public management have proceeded in tandem.

Formalizing the market sector does not ensure reciprocal changes in public institutions, however. Informality is as much a matter of culture as of practice; it defines social roles, relationships, and legitimate and expected behavior, and it persists even when the underlying conditions that gave rise to it vanish. There are quite a few countries in which the development of the public sector has not kept pace with economic changes. These countries typically have a competitive sector that is open, formal, and lightly regulated, as well as a heavily regulated sector that depends on informal contracts, embedded traditions, and government protection. The two cultures can operate independently of one another for an extended period, but sooner

or later they will be driven by scandal, financial mismanagement, or citizen pressure to modernize the public sector.

Second, modernizing the public sector means establishing reliable external controls, as described above. As old-fashioned as external controls may seem to be, they are building blocks for a formal, rule-based, honest public sector. Operating in an externally controlled environment is an essential phase in the development process. It gives managers the skills to manage on their own, builds trust between central controllers and line managers and confidence between citizens and government, and encourages managers to internalize a public ethic of proper behavior. As these basic conditions of formal management take root, it should be possible for central controllers to ease the regulations by giving line managers broader discretion in operating their programs.

This process, however, can bear fruit only if the controls are exercised in a fair and realistic manner. In the case of civil service rules, this means that pay levels rise as the economy develops, the number of ghost positions declines, and public employees are given opportunities to acquire new skills and advance professionally. If these conditions are absent, learning will take place, but it will be pathological: how to beat the system, how to outmaneuver the controllers, how to get paid without really working, and so on. Realism must also pervade budgeting, another arena that often is infected by pervasive informality. The budget presented to the parliament must be one that can be implemented, not a political wish list that promises more than the government intends to spend. Moreover, agencies must inculcate the habits and ethic of spending according to the plans laid out in the budget. In other words, the budget must be treated as an implicit contract. Only then does it make sense to convert the budget into an explicit contract.

Third, politicians and officials must concentrate on the basic process of public management. They must be able to control inputs before they are called upon to control outputs; they must be able to account for cash before they are asked to account for cost; they must abide by uniform rules before they are authorized to make their own rules; they must operate in integrated, centralized departments before being authorized to go it alone in autonomous agencies.

Once the basics have been mastered, the public sector should be organized according to the principles of internal control. External control and New Zealand-type managerial discretion are not the only options for organizing governmental operations. Internal control is a third possibility. In a formal sense, internal control refers to the systems and procedures used by agencies to assure compliance with rules and to safeguard public assets. In a behavioral sense, internal control means that the rules are accepted as fair, workable, and legitimate. Without this normative underpinning, no system of internal control can be effective.

In practice, internal control gives managers broader discretion; it shifts the focus from ex ante control to ex post audit, from control of individual actions to control within a broad band, from reviewing specific actions to reviewing systems. It means

for example, that civil service rules dictate the total number of positions or the total within broad employment categories and that operating departments make their own hiring decisions subject to oversight by central agencies. In the financial sphere it means that if funds are available, agencies can make purchases, authorize travel, and take other spending actions without obtaining prior approval.

Singapore illustrates the progression from external to internal controls and thence to New Zealand–type arrangements. On gaining independence and for many years afterward, Singapore had a line-item budget that specified the positions to be filled and the items to be purchased. During the 1980s block budgets were adopted that shifted the government from external to internal control, and in the mid-1990s a "budgeting for results" system was adopted that implements several elements of the New Zealand model.

Singapore provides another lesson for countries seeking to advance to the first rank of developing countries. The process of development does not have to stretch out over generations. If development proceeds in a logical order, progress can be rapid, especially if modernization of public institutions advances apace with modernization of the market sector. To many developing countries, New Zealand is at the cutting edge in public management, but they will not get there by taking shortcuts that turn into dead ends.

Notes

Allen Schick is a consultant in the Public Sector Group, Poverty Reduction and Economic Management, of the World Bank. He wrote this article while on assignment to the Economic Development Institute.

1. The term is taken from Davis, Sullivan, and Yeatman (1997). See especially Matheson (1997) in that volume.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Davis, Glyn, Barbara Sullivan, and Anna Yeatman, eds. 1997. *The New Contractualism?* South Melbourne: Macmillan Education Australia PTY Ltd.

de Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World.* New York: Harper and Row.

Matheson, Alex. 1997. "The Impact of Contracts on Public Management in New Zealand." In Davis, Sullivan, and Yeatman, eds. (1997), pp. 164–79.

Schick, Allen. 1996. "The Spirit of Reform: Managing the New Zealand State Sector in a Time of Change." Report prepared for the State Services Commission, Wellington, New Zealand. Processed.

Deforestation and Forest Land Use: A Comment

Jeffrey R. Vincent • Malcolm Gillis

Hyde, Amacher, and Magrath (1996) imply that deforestation and timber rents (logging revenue minus logging costs other than timber fees) are not subjects that justify policymakers' attention, arguing that market responses limit the scope of deforestation and that rents are usually small. But they fail to recognize that land markets will not develop efficiently, nor will efficient levels of forestry investments occur, when policy distortions and other factors obstruct the conversion of open-access forests to private or communal ownership. For these reasons rates of deforestation can be far above optimal levels. Contrary to the authors' claims, timber rents often (although not always) are large in developing countries. Moreover, the allocation of rents between loggers and the government owners of public forests can indeed affect the profitability of forestry (and thus deforestation), the intensity of timber harvesting, and national welfare.

"Deforestation and Forest Land Use" by Hyde, Amacher, and Magrath (1996) addresses one of the most prominent environmental issues of the last two decades: global deforestation. We agree with many of the article's main points, in particular that "market responses . . . create limits to potential deforestation" (p. 242). As the authors note, this is a modest argument. Many economists have written that, in theory at least, one would expect forest areas first to decline and then to stabilize in response to changes in the prices of land, labor, capital, and forest products (see, among others, Clawson 1979; Berck 1979; Sedjo and Lyon 1990; Vincent 1992; and Vincent and Binkley 1992). The historical pattern of land use in the mature market economies of North America, Europe, and East Asia illustrates this process nicely, as does the emerging pattern of land use in some rapidly industrializing countries such as Malaysia (Vincent and Hadi 1993). But because the story is an important one, it is worth retelling, especially with the fresh information contained in the section "The Empirical Evidence."

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 133–40

The authors push their argument too far, however, in three instances: first, in downplaying the potential inefficiencies associated with deforestation; second, in stating that "forest rent is seldom large" (p. 240); and third, in minimizing the economic importance of rent allocation between public owners and private users. In so doing they may have led readers to conclude that policymakers in developing countries and their advisers need not be concerned about deforestation or timber rents. In our view, such a conclusion is far off the mark.

Deforestation

The authors make several provocative assertions about deforestation. In particular they claim that capitalized forest values net of "the costs of establishing and protecting permanent rights to the land" are "generally" negative (p. 226), and that the capitalized values of "most remaining tropical forests" are negative even if one ignores the costs of establishing property rights (p. 227). They offer no empirical evidence to support these assertions. This is a critical omission. If the assertions are true, then it follows that the degradation and conversion of open-access tropical forests to other uses do not generate significant efficiency losses. In effect, the authors assert away the essential economic concern about deforestation.

They also implicitly assume that individuals, households, communities, and firms face few obstacles in converting land that does have a positive value in forestry uses from de jure or de facto open-access to private or communal ownership (pp. 225–7, 229). If that is true, then property rights and land markets might indeed emerge in an efficient manner over time, and deforestation might not be associated with significant inefficiencies. We agree that rural areas are much more marketized than is popularly supposed. But the efficiency of markets depends critically on the existence of well-defined, enforced property rights. In the case of plantation forestry investments in Chile, cited more than once by the authors to illustrate the presumed insignificance of policies relative to market adjustments, markets did not work their magic until the government assured potential investors that forests would not be nationalized (Vincent and Binkley 1992). In many other developing countries, including the world's two leading tropical timber producers, Indonesia and Malaysia, the ownership of forests by parties other than the state is (or was until recently) prohibited.

Without secure legal rights to forest land, economic agents are understandably reluctant to make the investments in forest management or forest plantations that are necessary for dynamically efficient forest-sector development. The authors dismiss such problems as "delays," stating that "there is no empirical evidence of their economic importance" (p. 236). But they provide no empirical evidence to demonstrate that institutional obstacles and the "delays" they generate are *not* important. Indeed, on pp. 238–39 they provide several examples of policy distortions that prob

ably do generate significant inefficiencies. Others are outlined in one of the articles they cite (Gillis 1988a).

Surely governments ought to be concerned about removing obstacles to the development of efficient land markets (see Repetto and Gillis 1988, Vincent and Binkley 1992). The observation that deforestation might not "attain its physical limit" (p. 223) is not very reassuring from a social welfare standpoint: the cumulative amount and rate of deforestation might still be far above socially efficient levels.

The Magnitude of Timber Rents

Timber rent is the return over and above the factor costs of logging: the difference between total revenue from timber sales and total costs of harvesting and delivery (including a normal profit margin, but excluding the fees charged by the government in the form of taxes, licenses, and so forth). In light of evidence from the literature and from the field, the authors' claim that "forest rent is seldom large" (p. 240) is astonishing. Many studies conducted since the 1970s have documented the existence of large rents in the virgin or semi-virgin (logged over, but with residual virgin trees) forests that have been the locus of most commercial logging in developing countries (Page, Pearson, and Leland 1976; Ruzicka 1979; Gillis 1980; Repetto and Gillis 1988; Vincent 1990; Sizer and Rice 1995; Sizer 1996). The authors cite almost none of this literature. In fact, they appear convinced that such rents simply cannot exist, commenting that "Surely all private operators would have had sufficient incentive to harvest each year all the way to the geographic point where their access and harvest costs depleted the entire value of the standing resource" (p. 240).

There are several reasons why large rents can, and indeed do, exist in many tropical forests. The first comes from basic theory. An old-growth timber stock is like a deposit of a nonrenewable resource. When a nonrenewable resource is depleted at an optimal rate, its marginal rent (the rent on the last unit extracted in a given period) rises over time at a rate equal to the discount rate (the return on an alternative investment). This reflects the rising cost of depletion as the resource becomes more scarce, and it provides the fundamental economic reason why a rational user concerned with future as well as present returns will not harvest "all the way to the geographic point. . . ." The only difference in the case of forestry is that the rate of increase in the marginal rent (which forest economists call stumpage values) diminishes as logging proceeds and more vigorously growing, second-growth forests, which offer a biological rate of return because of growth, replace old-growth forests, which have no net timber growth.

Contrary to the authors' claims, long-run data on stumpage values do exist (for instance, see the sources in Binkley and Vincent 1988 and Nordhaus 1992), and they are broadly consistent with this theoretical story. For example, softwood stumpage

values in the U.S. South rose in real terms by 4.6 percent a year from 1910 to 1943 and 3.1 percent a year from 1943 to 1985; they are predicted to rise at around 2 percent a year into the early 2000s (Binkley and Vincent 1988). Should one be surprised that rents for tropical timber, whose products compete with ones made from temperate timber, are high today after a century of steadily rising values for temperate timber?

This theoretical explanation for large tropical timber rents is not entirely convincing, however. The absence of secure property rights in many developing countries implies that loggers do not make socially efficient decisions: in comparing the returns to current and future harvests, they place too little weight on the latter. In the extreme, if loggers completely ignore depletion costs, then harvesting will indeed proceed "all the way to the geographic point . . ." where marginal rent equals zero. If that is the case, where do large rents come from?

The first potential source is from timber *within* the margin: timber in forests with easy terrain and near roads, mills, and ports. Indeed, this "inframarginal" rent was the major source of the large timber rents in Malaysia estimated by Vincent (1990). The possibility of large rents within the margin coexisting with rents equaling zero at the margin has been known since the time of classical economists such as Ricardo and von Thunen.

The second potential source of large rents is from government regulation of logging activity. Logging in Southeast Asia, the world's major source of tropical hardwood logs and products, is not completely uncontrolled. State and federal governments in Malaysia and Indonesia control access and utilization of the forest, although not perfectly, by issuing harvesting licenses, and by setting cutting limits at some minimum diameter. Such regulations prevent loggers from harvesting "all the way to the geographic point . . ." and result in positive rents at the harvesting margin.

Third, the expansion of infrastructure (logging roads, ports, and so forth), the development of lower-cost logging technologies, and increases in the number of species and sizes of roundwood accepted by markets have combined to increase rents over time. Such changes, which are typically not fully anticipated by timber markets, make rents available for loggers, concessionaires, and their political patrons. The authors downplay these changes as "specialized cases" of a return to "a relaxation of a restriction on forestry activity" (p. 240). Such changes do indeed relax restrictions on forestry activity, but they are far from occasional curiosities. Indeed, they have been the driving forces of the expansion of the tropical timber industry since the end of World War II.[1] Today, they are causing a historic shift of the industry from Southeast Asia to Latin America. It is unlikely that timber barons in Kuching and Jakarta will abandon their pursuit of concessions in Surinam, Guyana, and the Amazon after reading in Hyde, Amacher, and Magrath that tropical timber rents are no more than "relaxations of restrictions." Paraphrasing Gertrude Stein, a rent is a rent is a rent. Nor will Malaysian logging companies stop bidding

as much as 40,000 ringgit a hectare ($16,000) for harvesting rights in peninsular Malaysia (Awang Noor 1996).

In sum, there is no mystery why large rents are both theoretically possible and do in fact exist in many tropical timber-producing countries. Some of the evidence the authors present to deny this situation simply does not withstand scrutiny. For example, they imply that rents were low in the Philippines in the 1980s by juxtaposing the example of the Philippines Bureau of Forestry Development, which lost money on its timber sales in the early 1980s (revenues from timber fees were less than the bureau's expenses), with that of below-cost timber sales in the United States. But timber fees were low in the Philippines not because rents were low, but because the bureau captured less than 10 percent of the available rents. Low rent capture, not low rent, was in fact the main point of the article by Boada (1988) that they cite.

The authors draw the forestry rent gradient in their figure 1 so that it "shows that" (p. 226) the value of forest land is low, but they could equally well have drawn it higher. The articles they cite lead us to speculate that the developing countries they had in mind in drawing this figure are ones whose commercial forests are a small proportion of total forests, such as China, India, Malawi, Nepal, Pakistan, and the Philippines. It is not surprising that average forest rents are low in such countries, which, because of ecological factors (low rainfall, say) or historical reasons (when considerable depletion has already occurred), do not have heavily stocked forests. But one should be wary of extrapolating from experience in such countries to the entire developing world, just as one should be wary of extrapolating from deserts to the global value of commercial agricultural land.

Allocation of Timber Rents

By contrasting two of our papers (Gillis 1988a; Vincent 1990) against two others (Paris and Ruzicka 1991; Hyde and Sedjo 1992), Hyde, Amacher, and Magrath imply that we have argued that timber rents in public forests in developing countries are always large and always belong to the government, and that the assignment of these rents always has an efficiency impact. This is not an accurate representation of our work. Regarding the first claim, in Gillis (1988a) and Vincent (1990), we emphasized that rent can vary substantially from stand to stand. Gillis (1988a; see also Gillis 1980) dealt explicitly with variations in rent within and between stands. As noted earlier, the model in Vincent (1990) was based on the existence of rent variation within the margin of utilization. Nor have we claimed that large rents in virgin forests imply that rents in second-growth forests, which are what matter for long-run land-use decisions, are also large. Vincent (1990) examined in great detail how the present value of net returns from second-growth harvests varies according to differences in prices, costs, rotation lengths, growth rates, and discount rates.

As for the second and third claims, we have never stated that the complete capture of resource rent by government agencies is necessary for efficient land-use allocation or efficient levels of forest management investments. Vincent (1993) cited quotations from our work emphasizing that many aspects of rent capture involve primarily distributional issues.

We, and others before us (Page, Pearson, and Leland 1976; Schmithüsen 1976; Ruzicka 1979), *have* argued that the fees that forestry agencies levy to capture timber rents can affect marginal harvesting decisions and result in inefficient harvest levels if they are not set appropriately. This follows from basic concepts of resource taxation. Hyde and Sedjo (1992) accepted the validity of this point, although as noted in Vincent (1993), they misinterpreted much of the analysis in Gillis (1980, 1988a) and Vincent (1990).

Hyde, Amacher, and Magrath ignore this and several other potential links between government rent capture and efficiency within the forest sector. For example, Page, Pearson, and Leland (1976) pointed out that a government's failure to set timber fees at levels comparable to competitive stumpage values can destroy rents by artificially maintaining the viability of inefficient logging firms. (Gillis 1988a showed how industrial policies—high levels of protection to Indonesian plywood mills—destroyed rents by promoting inefficient local processing.) That timber fees can affect land-use decisions is evident from figure 1 in Hyde, Amacher, and Magrath: timber fees reduce the private return to forestry, shift the forest rent gradient downward, and thus reduce equilibrium forest area. Whether this effect is good or bad depends on the relative social returns to different land uses.

The allocation of rents from logging also has important implications beyond sectoral efficiency. One would expect national welfare to be enhanced if rents are retained within the country instead of being repatriated by transnational logging companies. Concerns about retained value triggered early studies on rent capture in the tropical timber sector, such as those by Page, Pearson, and Leland (1976), Ruzicka (1979), and Gillis (1980). The authors cite none of those works. We seriously doubt that Papua New Guinea and the Latin American countries that are the new frontiers for tropical logging will be better off if they grant harvesting rights to Asian logging companies for a song.

The authors overlook a second link to national welfare: the potential efficiency gains that result from the state's ability to capture a higher proportion of timber rents and to reduce existing taxes that are more distortionary. If the authors doubt the significance of such measures, they should consider how low rates of rent capture in natural resource sectors have undermined economic recovery in Russia by forcing the state to impose crushing taxes on income, commercial transactions, and international trade. Indonesia and Malaysia, at the other extreme, have been able to keep general tax rates low because they generate revenue from petroleum (though not timber) rents.

The authors claim (note 4) that analyses of the impacts of general economic policies upon the forest sector are scant. In fact, several such analyses have long been available. The book *Public Policies and the Misuse of Forest Resources* (Repetto and Gillis 1988) not only highlighted failures in forestry policies, but it also discussed the effects of macroeconomic, trade, monetary, and agricultural policies on the forest sector. Indeed, Gillis (1988a), which Hyde, Amacher, and Magrath cite, stressed these effects. Another chapter in the same volume depicted these effects in some detail for Ghana and several other African countries (Gillis 1988b). Xie, Vincent, and Panayotou (1997) cited nearly 20 computable general equilibrium modeling studies showing the impacts of general economic policies on the forest sector.

In conclusion, it will be unfortunate if readers of the *Research Observer* conclude from Hyde, Amacher, and Magrath that deforestation and timber rents are not worthy of policymakers' attention. Both are associated with policy issues whose resolution could have important impacts on the efficiency of land use, forest managements, and economic growth in developing countries.

Notes

Jeffrey R. Vincent is a fellow at the Harvard Institute for International Development. Malcolm Gillis is president of Rice University.

1. And such changes, not timber growth (p. 225), are the main causes of pulse-harvesting in the tropics.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Awang Noor Abd Ghani. Faculty of Forestry, Universiti Pertanian Malaysia. Personal communication, December 17, 1996

Berck, Peter. 1979. "The Economics of Timber: A Renewable Resource in the Long Run." *Bell Journal of Economics* 10(2) 447–62.

Binkley, Clark S., and Jeffrey R. Vincent. 1988. "Timber Prices in the U.S. South: Past Trends and Outlook for the Future." *Southern Journal of Applied Forestry* 12(1):15–18.

Boada, Eufresina L. 1988. "Incentive Policies and Forest Use in the Philippines." In Repetto and Gillis (1988).

Clawson, Marian. 1979. "Forestry in the Long Sweep of American History." *Science* 204.1168–74.

Gillis, Malcolm. 1980. "Fiscal and Financial Issues in Tropical Hardwood Concessions." Development Discussion Paper 110. Harvard Institute for International Development, Cambridge, Mass. Processed.

----. 1988a. "Indonesia: Public Policies, Resource Management, and the Tropical Forest." In Repetto and Gillis (1988).

———. 1988b. "West Africa: Resource Management Policies and the Tropical Forest." In Repetto and Gillis (1988).

Hyde, William F., Gregory S. Amacher, and William Magrath. 1996. "Deforestation and Forest Land Use: Theory, Evidence, and Policy Implications." *The World Bank Research Observer* 11(2):223–48.

Hyde, William F., and Roger A. Sedjo. 1992. "Managing Tropical Forests: Reflections on the Rent Distribution Discussion." *Land Economics* 68(3):343–50.

Nordhaus, William D. 1992. "Lethal Model 2: The Limits to Growth Revisited." *Brookings Papers on Economic Activity* 2(1):1–43.

Page, John M. Jr., Scott R. Pearson, and H. E. Leland. 1976. "Capturing Economic Rent from Ghanaian Timber." *Food Research Institute Studies* 15(1):25–51.

Paris, Remy, and Ivan Ruzicka. 1991. "Barking Up the Wrong Tree: The Role of Rent Appropriation in Sustainable Tropical Forest Management." Environment Office Occasional Paper 1. Asian Development Bank, Manila. Processed.

Repetto, Robert, and Malcolm Gillis, eds. 1988. *Public Policy and the Misuse of Forest Resources.* Cambridge, U.K.: Cambridge University Press.

Ruzicka, Ivan. 1979. "Rent Appropriation in Indonesian Logging: East Kalimantan, 1972/73–1976/77." *Bulletin of Indonesian Economic Studies* 15(2):45–74.

Schmithüsen, Franz. 1976. "Forest Utilization Contracts on Public Land in the Tropics." *Unasylva* 28(1):52–73.

Sedjo, Roger A., and Kenneth S. Lyon. 1990. *The Long-Term Adequacy of World Timber Supply.* Washington, D.C.: Resources for the Future.

Sizer, Nigel. 1996. *Profit without Plunder: Reaping Revenue from Guyana's Tropical Forests without Destroying Them.* Washington, D.C.: World Resources Institute.

Sizer, Nigel, and Richard Rice. 1995. *Backs to the Wall in Suriname: Forest Policy in a Country in Crisis.* Washington, D.C.: World Resources Institute.

Vincent, Jeffrey R. 1990. "Rent Capture and the Feasibility of Tropical Forest Management." *Land Economics* 66(2):212–23.

———. 1992. "The Tropical Timber Trade and Sustainable Development." *Science* 256(5064):1651–55.

———. 1993. "Managing Tropical Forests: Comment." *Land Economics* 69(3):313–18.

Vincent, Jeffrey R., and Clark S. Binkley. 1992. "Forest-Based Industrialization: A Dynamic Perspective." In Narendra P. Sharma, ed., *Managing the World's Forests.* Dubuque, Iowa: Kendall-Hunt.

Vincent, Jeffrey R., and Yusuf Hadi. 1993. "Malaysia." In National Research Council, *Sustainable Agriculture and the Environment in the Humid Tropics.* Washington, D.C.: National Academy Press.

Xie, Jian, Jeffrey R. Vincent, and Theodore Panayotou. 1997. "Computable General Equilibrium Models and the Analysis of Policy Spillovers in the Forest Sector." Environment Discussion Paper 19. Harvard Institute for International Development, Cambridge, Mass. Processed.

Deforestation and Forest Land Use: A Reply

William F. Hyde

In our recent article (Hyde, Amacher, and Magrath 1996), we argued that the "topic of deforestation is seldom examined from the perspective of prices and responses to resource scarcity" (p. 223). Our point was not that the market is perfect but rather that market responses limit the extent of deforestation—even in regions of attenuated property rights or open-access forest resources. The market sets this limit because as deforestation proceeds, the increasing scarcity of the resource induces rising prices, which in turn induce private investments to satisfy both commercial and subsistence household uses of forest resources. As a result, deforestation proceeds to the geographic points where the harvest and access costs associated with removing the remaining natural stock equal the costs of reforestation on land that has established and enforceable property rights. This means that a considerable area of the world's natural forest is not endangered.

Therefore, we argued that concern for the total forest resource and its protection is misguided. It would be better to concentrate on the price increases induced by deforestation and the costs of private reforestation with an eye to identifying the level of prices and costs that will alter the boundaries of several important classes of forest land, and on concentrations of nonmarket, forest-based environmental services and public policy failures that distort market incentives.

Vincent and Gillis, in their article in this issue, feel that we err on three counts: deforestation, the magnitude of timber rent, and rent allocation. I will address these in order.

Deforestation

Vincent-Gillis suggest that we diminish the inefficiencies associated with deforestation. The source of our difference is their focus on our "Simple Model" of market

The World Bank Research Observer, vol. 13, no. 1 (February 1998), pp. 141–45

forest values (pp. 225–30). We qualified the model with a more extended discussion of nonmarket factors (pp. 234–36) and policy failures (pp. 236–42).

Vincent-Gillis are concerned with our argument that net market forest values are generally negative. They feel that we deny the essential economic concern with deforestation. But surely some forest lies beyond the reach of any active harvest activity. Some even lies beyond the reach of open-access activity. Surely, this is land with a negative net value in the local market. A recent report from the World Resources Institute (Bryant, Nielson, and Tangley 1997) suggests that 40 percent of the world's forests lie beyond the reach of market activity. Table 1 (p. 230) reports estimates that are at least this large. In addition, there is a region of the world's forest where previous harvests of the natural stock have left standing trees that are either too immature or too sparse to have positive net value. In most time periods, this land also has a negative net market value. Vincent himself was the first to my knowledge to use the term "pulse" harvests (Vincent and Binkley 1992) to identify this region.

These observations, however, do not deny the essential economic concern about deforestation. My own impression is that this concern has to do with nonmarket values (global change, biodiversity, erosion, or watershed protection), especially those attached to the remaining stock of natural forest. In examining these values, we argued that global concern for their protection should be concentrated in those higher market-value forests where nonmarket values are at risk. We find this approach preferable to wasting environmental policy on regions that are not accessible in any case and where the nonmarket environmental values are protected without policy intervention.

Vincent-Gillis state that we "implicitly assume that individuals and firms face few obstacles in converting (open-access) land" to private property (p. 134). But we explicitly discussed the problems of establishing secure tenure and cited some of the recent literature on pp. 238–39. Vincent-Gillis's arguments and ours are similar.

Contrary to Vincent-Gillis's claim, we did not dismiss the problems of establishing secure tenure as a market delay. The "delay" they speak of (p. 236) specifically referred to our simple model and the response by investors to rising market prices We observed that there is no empirical evidence on the magnitude of this particular market lag. This does not mean that the lag does not exist, or that it is the only identifiable market failure, but only that the lag and its effects have not been measured successfully. As for institutional obstacles to forest investment, we discussed macroeconomic policy (p. 237), legal institutions (p. 238), spillovers from agricultural policy (p. 238), and public forestry agencies (pp. 239–40). In each case, we illustrated the problems these pose for private forestry activities, and either the incentives created for deforestation or the obstacles for reforestation. We also discussed publicly funded research, technology transfers, and research to enhance forest production and to identify improved institutional arrangements (pp. 241–42). Contrary to Vincent-Gillis's claim, we never presume the "insignificance of policies rela-

tive to market adjustments" (p. 134). I would say that we stressed the importance of these policies in modifying market adjustments and that our conclusion specifically encouraged the assessment of their impacts on the boundaries of the various classes of forest land.

The Magnitude of Forest Rent

Vincent-Gillis are concerned that we underestimate the importance of timber rents. Because our article was about deforestation and because the concern with deforestation is comprehensive, our context is the entire global forest, while Vincent-Gillis's context is with the share of mature natural forests where commercial timber harvests are important.

Bryant, Nielson, and Tangley (1997) suggest that the share of mature natural forest that is subject to commercial timber harvest and that may have a positive timber rent is 28 percent of the world's total forest. We agreed that concerns with timber rent are valid in this region (p. 238), citing the primary literature as well as examples from our own experience to illustrate that these rents can be large (p. 239). These examples raise questions about the prospective occurrence of large rents in other cases. We think that the important questions have to do with what caused the large rents to arise and whether these sources of rent are instructive for policy.

The mature natural forests that are the sources of commercial harvests and the common sources of timber rents are generally at or near the margins of previous economic harvest activity—otherwise they would have been harvested earlier. Gradual long-term price increases do give rise to rents at these locations, but they are unlikely to induce large rents before harvesting extends into these regions. Large rents must be caused by sharp market changes or by something that constrained gradual harvest expansion. The former are uncommon. The two obvious candidates for the latter are a change in infrastructure and successful enforcement of some government logging restriction, as Vincent-Gillis point out. Most of our discussion of rent was devoted to the same two sources, particularly changes in access (for example, new roads) and the particular biological formulation of the "allowable cut" harvest policy that is common in one form or another to most government forestry agencies.

These observations still "beg questions about the management activities or policies that created the rent. Does the same policy exist elsewhere? Could more rents be captured by applying the same management activity or the same change in forest policy to other timberstands . . . ?" (p. 240) And, in our context, what would be the impacts on deforestation and on incentives to reforest?

Finally, and contrary to Vincent-Gillis, we did not deny the existence of long-run stumpage price series. Rather we noted that these data are "not reliable" for purposes of prediction (p. 231). They are unreliable because stumpage value is the residual left

after all harvest, transport, and processing costs are deducted. This residual varies greatly, even within a region and often within a harvest site, according to differences in wood quality and location, and on-site harvest difficulties. Therefore, stumpage values from one time or place are not good predictors of stumpage values at another time or place, and composite stumpage price series from one time are not good predictors of composite stumpage prices in another time. This is one reason why Barnett and Morse (1963) used lumber prices, not stumpage prices, in their classic analysis of intertemporal trends in resource prices in the United States.[1]

Rent Allocation

Vincent-Gillis believe that we minimize the importance of rent allocation, and that we imply they believe three things are always true. We would never say someone else "always" believes anything. To clarify any possible misunderstanding I would restate our sentence (p. 239) to read as follows: "Vincent-Gillis have emphasized cases of large timber rent, and they have been concerned with claiming a larger share of this rent for the government. Vincent has also argued that rent allocation has an effect on efficient output levels." I think the articles Vincent-Gillis cite in their comment support this revised statement. Nevertheless, the restatement does not alter our point that all discussions of timber rent should lead to questions about the further occurrences of these rents, the policies that give rise to them, and the application of similar policies in other timberstands and forests.

As for the merit of rent capture by public agencies, the difference between Vincent-Gillis's position and ours is that we believe that allocation is a distributive matter of undetermined merit. That is, it is an empirical question that must be judged by the performance of the government agency or forest concessionaires receiving the rent in each particular case. We have a preference for the most socially productive use of the rent—rather than for a favored agent.[2]

Of course, policy can modify and even destroy rents, as Vincent-Gillis argue and we agree. Administrative behavior can also modify or destroy rent. We all know of government agencies as well as private operations that dissipate rents with poor decisions, just as we know of those that collect and reinvest rents in ways that improve the general welfare.[3] This is what makes the preferred allocation a distributive issue whose resolution depends on the performance of the agents involved.

In conclusion, our concern was with global deforestation, whether due to commercial forestry, agricultural land conversion, or subsistence household uses of the forest. Our primary argument was that broad-based forest policy interventions designed as general protection against world deforestation are not well advised because specific markets for the various products and services of the forest already exist. These specialized markets divide forests into at least three distinct land use

classes: sustainable private forests and forested commons; unsustainable open-access forests near the frontier; and an unused residual forest. Forest policy will affect each of these land categories differently. Therefore, it would be better to address the market and policy failures that shift the boundaries of one or another of these classes of land use, and to anticipate that most policies will selectively affect some, but not all, of these boundaries. I think our original article was clear on these points.

Notes

William F. Hyde is professor in the Forestry Department at Virginia Polytechnic and State University, Blacksburg, Virginia.

1. Smith (1979) raised doubt about the statistical reliability of many primary resource prices for prediction. Jung, Krutilla, and Boyd (forthcoming) demonstrate the unreliability of composite stumpage price series.

2. Contrary to Vincent-Gillis, allocation does not alter the efficient level of production in our figure 1. Also contrary to Vincent-Gillis, rents are not low because of the way we drew the rent gradient. The forest rent gradient almost always starts below the agricultural land value gradient and extends to its right. Forest rents are low because forests are lower-valued uses of the land. A steeper gradient in our figure 1 would not alter these basic points.

3. Repetto and Gillis (1988), who are cited by Vincent-Gillis, give many examples of government actions that dissipated timber rents or government agencies that did not have the administrative ability to capture some rents. These examples would seem to urge caution on any view that government is a generally preferred collector of rents.

References

Barnett, Harold J., and Chandler Morse. 1963. *Scarcity and Growth. The Economics of Natural Resource Availability.* Baltimore, Md.: Johns Hopkins University Press.

Bryant, Dirk, Daniel Nielson, and Laura Tangley. 1997. *The Last Frontier Forests. Ecosystems and Economies on the Edge.* Washington, D.C.: World Resources Institute.

Hyde, William F., Gregory S. Amacher, and William Magrath. 1996. "Deforestation and Forest Land Use: Theory, Evidence, and Policy Implications." *The World Bank Research Observer* 11(2):223–48.

Jung, Chulho, Kerry Krutilla, and Roy Boyd. Forthcoming. "Aggregation Bias in Natural Resource Price Composites: The Forestry Case." *Energy and Resource Economics.*

Repetto, Robert, and Malcolm Gillis, eds. 1988. *Public Policy and the Misuse of Forest Resources.* Cambridge, U.K.: Cambridge University Press.

Smith, V. Kerry. 1979. "Natural Resource Scarcity: A Statistical Analysis." *Review of Economics and Statistics* 61:423–27.

Vincent, Jeffrey R., and Clark S. Binkley. 1992. "Forest-Based Industrialization: A Dynamic Perspective." In Narendra P. Sharma, ed., *Managing the World's Forests.* Dubuque, Iowa: Kendall-Hunt.

Vincent, Jeffrey R., and Malcolm Gillis. 1997. "Deforestation and Forest Land Use: Comment." *The World Bank Research Observer* 13(1):133–40.

Promoting Efficient Rural Financial Intermediation

Jacob Yaron • McDonald Benjamin • Stephanie Charitonenko

Although governments have traditionally used subsidized credit programs to promote agricultural growth, this approach has generally failed to improve incomes and alleviate poverty in rural areas. It has also led to the mistaken belief that rural credit programs cannot be profitable. A new approach seeks to raise standards of living in rural areas by casting the government in a very different role—one of setting a favorable legal and policy environment for rural financial markets and addressing specific market failures cost effectively through well-designed and self-sustaining interventions. There is evidence that this approach can be highly successful. The Village Bank system of Bank Rakyat Indonesia has shown that financial services can be extended to millions of low-income rural clients without relying on subsidies. Indeed, the program has generated enormous profits for the bank by using simple, innovative, and largely replicable techniques.

Providing affordable financial services to the rural population has been an important component of development strategy for the last several decades. Direct interventions in rural financial markets to stimulate growth and reduce poverty—through a blend of targeted credit programs, interest subsidies, and other government policies—became widespread in the 1950s, when Keynesian economics inspired many governments to design fiscal interventions at the macroeconomic level. But these direct interventions have generally been disappointing and have tended to retard, rather than promote, the development of financial services in rural areas. One explanation is that these policies were based on serious misconceptions about the real challenges facing rural communities and were directed more toward the symptoms rather than the causes of inadequate rural financial intermediation.

More recent developments in the provision of rural financial services (both savings and credit) have demonstrated that proper institutional design and adherence to appropriate policies pay off handsomely and have the potential to generate substantial achievements in terms of both sustainability and greater institutional outreach.

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 147–70

Although the new approach focuses on the same objectives, that is, income expansion and poverty reduction, the perceived challenges and ways of addressing them are strikingly different. The principal change involves promoting deep and efficient rural financial markets by creating a favorable policy environment, improving the legal and regulatory framework that supports rural financial markets, and addressing specific market failures in cost-effective ways through well-designed, self-sustaining interventions.

The Traditional Approach to Rural Finance

Throughout the world, governments have intervened extensively in financial markets in general and rural financial markets in particular. The extent of intervention has varied from indirect measures aimed at improving the policy environment (for example, by addressing incentive problems and regulating financial intermediaries), to direct steps to increase or supplant credit provided by private lenders. Many countries eager to channel funds to farmers directed private banks to make concessional loans to agriculture, or they established and supported state-owned agricultural credit institutions.

Traditionally, the case for subsidized agricultural credit programs has been based on the following arguments: governments should focus on agriculture to promote rural development; agriculture is undercapitalized; farmers need cheap credit to encourage them to adopt modern technology and to compensate them for policies that are biased in favor of urban dwellers; farmers are too poor to save; and private banks provide little or no credit, forcing small borrowers to use moneylenders who charge usurious interest rates. Donors provided considerable support for subsidized credit; the World Bank, for example, lent $16.5 billion in agricultural credit under largely traditional programs prior to 1992 (World Bank 1993).

These programs have generally had a limited outreach and resulted in huge costs, with little identifiable impact at the farm level. In an extreme example, during the 1980s, one Latin American rural financial institution with more than 500 branches and 27,000 employees received $10.3 billion in fiscal and quasi-fiscal transfers (that is, capital injections and interest subsidies), while recovering only 10–15 percent on its portfolio and serving only 2 percent of the rural population. Elsewhere, government-sponsored rural credit programs and institutions from Peru to Malawi to Indonesia have collapsed under the weight of losses generated by traditional directed credit strategies.

These failures are largely explained by the pursuit of short-term objectives framed in terms of agricultural production gains rather than long-term objectives aimed at the sustained expansion of rural incomes. The excessive focus on disbursing cheap

agricultural credit has typically resulted in programs with a poor credit culture, manifested by a dependency on subsidies, low recovery rates, inadequately diversified portfolios, mistargeting of credit (Khan 1977), and rent-seeking by credit officials and influential farmers (Ladman and Tinnermeier 1984). The tremendous potential for rural savings has also been neglected, and private for-profit financial institutions have been crowded out by state-owned rural financial institutions dependent on government subsidies.

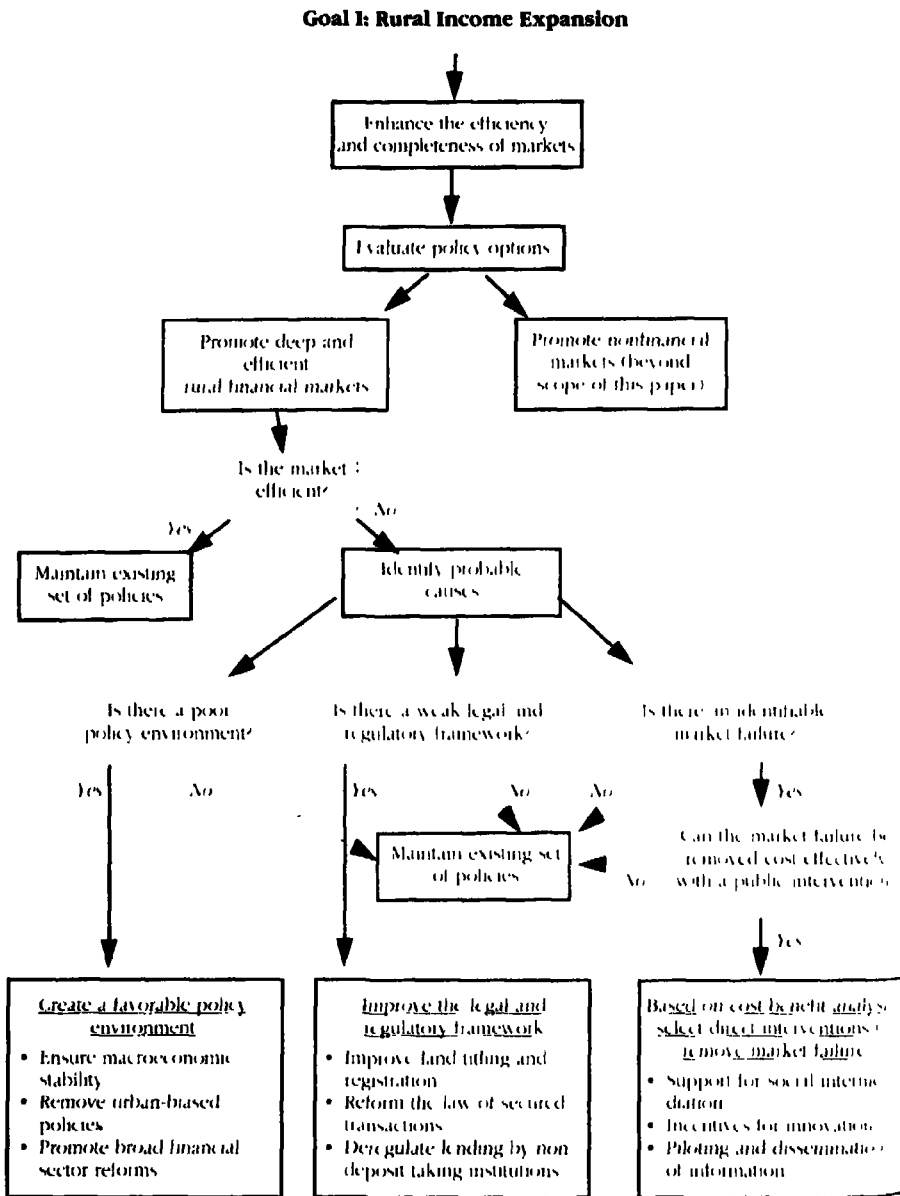
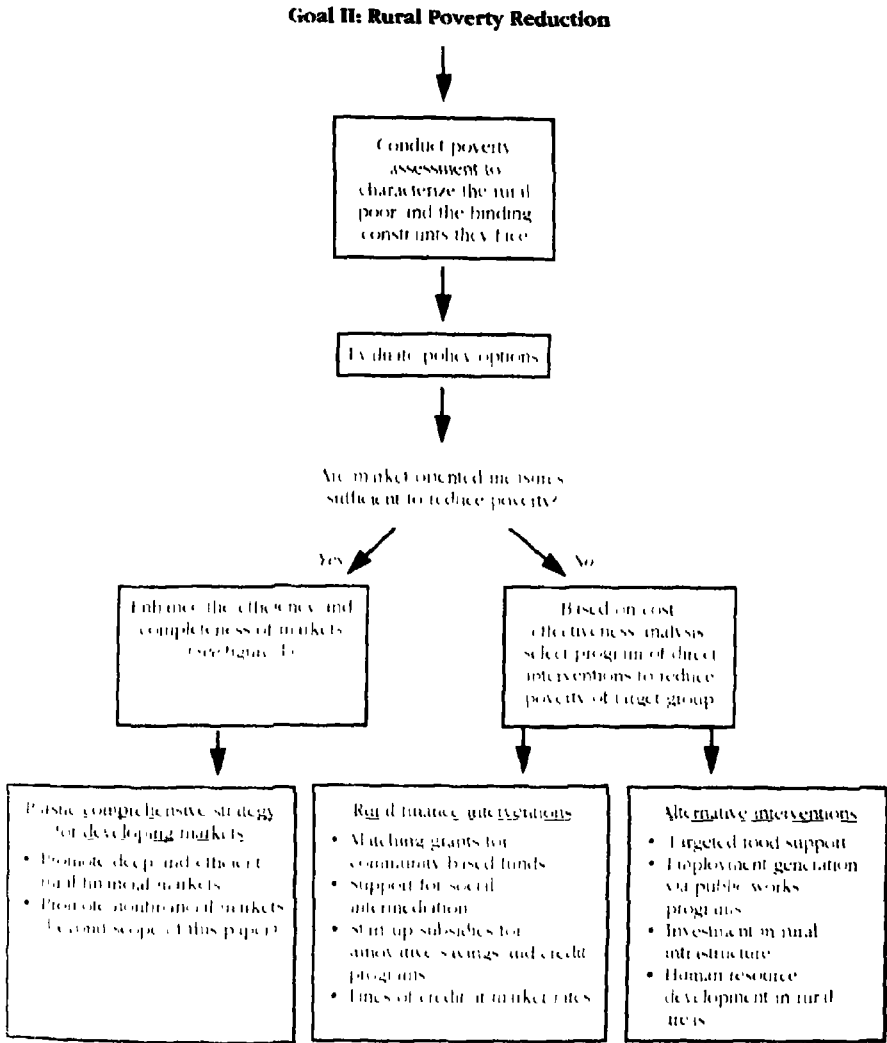
Although directed credit has been heavily criticized (Von Pischke, Adams, and Donald 1983), and more market-friendly approaches have been proposed for some time, many countries have resisted changing the rules under which state-owned financial institutions operate. Nevertheless, major reforms of rural credit systems have been launched in several countries, including India and Mexico, to ensure that public resources are used more effectively, to support the expansion of rural incomes, and to reduce poverty.

The New Approach to Rural Finance

The new approach continues to focus on income expansion and poverty reduction but makes the case for cost-effective alternatives, such as increased investment in rural infrastructure or in human development, to reach these goals. (For the literature on changes in rural finance, see Adams, Graham, and Von Pischke 1984; David and Meyer 1984; Gonzalez-Vega 1984; and Vogel 1984). Advocates of this approach propose that governments concentrate on establishing a favorable policy environment that facilitates the smooth functioning of rural financial markets while playing a more limited and efficient role in the direct provision of rural financial services. The factors that prevent rural financial markets from operating efficiently are recognized to be broader and include macroeconomic policies, weakly regulated financial sectors, institutional features (legal and regulatory), and specific constraints related to intermediation in rural areas. This approach sees the government's main task as creating a conducive environment for private intermediaries in rural financial markets (figures 1 and 2). The case for direct interventions depends on whether the objective is general rural income expansion or targeted poverty reduction.

Creating a Conducive Policy Environment

The starting point for formulating policies aimed at increasing rural incomes and reducing poverty is an assessment of the efficiency of markets, particularly rural financial markets, and of the causes of market inefficiencies. Typically, there are weak-

Figure 1. *Decision Tree to Promote Effective Financial Markets*

Source Yaron, Benjamin, and Piprek (1997)

nesses in the policy environment that hamper the development of rural markets, including financial markets. For example:

- Unsound macroeconomic policies result in volatility and high real interest rates that can adversely affect all financial intermediaries, while misaligned exchange rates distort price signals and lead financial markets to channel excessive resources to inefficient sectors.
- Development policies biased toward urban areas reduce the profitability of agriculture and nonfarm rural enterprises and devastate rural financial markets. Countries with the highest degree of discrimination against agriculture have had the lowest rates of economic growth (Schiff and Valdés 1992).
- Inadequate regulatory oversight, inappropriate interventions in financial markets, and financial repression increase the risks and constrain the development of financial markets.

Governments can promote financial markets in general by strengthening the supervision and prudential regulation of financial institutions, deregulating interest rates, reducing excessively high reserve requirements, and relaxing credit controls. Governments can also adjust the regulatory framework to facilitate operations in rural areas by community-based, deposit-taking intermediaries. Such policies would combine lower capital requirements with higher capital asset ratios and more circumscribed permitted activities to minimize regulatory arbitrage (see Berenbach and Churchill 1997 for similar regulatory issues in microfinance).

The Legal and Regulatory Framework

Too often the institutional foundations for financial markets in rural areas are absent. Lenders need a system that provides formal procedures for claims against property and enforcement of financial contracts. The more uncertain and expensive this process, the less willing are lenders to lend (Fleisig and de la Peña 1996). In many countries deficiencies in laws, regulations, and institutions prevent the formal sector from delivering credit to farmers, rural businesses or even nonbank creditors (typically traders), who have many advantages in efficiently reaching poor rural borrowers.

The required changes needed to expand access to credit in rural areas include titling and registering land; reforming the law of secured transactions, such as legally acceptable forms of collateral; establishing legal registries and expanding the scope for private operation; lowering the costs of registration and foreclosure; drafting specific, clear, and limited homestead provisions; and removing interest rate ceilings. Well-designed programs to reform the laws of secured transactions have increased

the supply of credit and lowered interest rates, producing gains over time that have been estimated at several percentage points of gross domestic product GDP (Fleisig and de la Peña 1996). The costs of implementing such legal reform programs are usually remarkably low.

Designing and Justifying Direct Interventions in Rural Financial Markets

What role should government play in rural finance and development? There is growing recognition that governments should first and foremost facilitate the workings of the market so that private participants can allocate resources efficiently in response to price and profit signals.

Risks of Market Failure

At the same time, markets may fail for several reasons, because the assumptions that are required in theory for efficient market-based resource allocation may not hold in practice. For example, individuals may not bear the full benefits or costs of their actions. Or externalities may arise because investors cannot capture the full benefits of their investments if they cannot exclude others from free-riding. Alternatively, individuals may fail to take into account the costs they impose on others when undertaking a given activity. A market may not have sufficient buyers and sellers or permit sufficient ease of entry and exit to ensure an efficient allocation of resources. In many countries market participants may not be able to enter easily into enforceable contracts. Finally, market participants may not be able to ensure against certain contingencies, although efficient markets for pricing and exchanging risks are required for optimal resource allocations when there is uncertainty. Rural financial markets, particularly in developing economies, generally have these shortcomings.

The critical factor that explains the externalities, missing markets, and local (competitive) monopolies in rural financial markets is imperfect information (Stiglitz 1996; Virmani 1982). Financial transactions in a given currency by their very nature involve a contractual exchange of cash for a promise of a future stream of payments, rather than a simultaneous exchange of cash or goods—or both—for goods. The promissory feature of financial transactions makes it essential for participants to be well informed about their counterparts' ability and willingness to honor contractual obligations. The absence of such information will constrain a lender's (or depositor's) ability not only to discern the creditworthiness of potential borrowers (or banks), but also to enforce contracts. These constraints point to an important role for government in regulating financial intermediaries, for in-

stance to limit excessive risk-taking by banks using other people's money, an
providing a sound legal and regulatory framework for enforcing contracts.

Rethinking Direct Interventions

Information constraints do not immediately justify direct government interventi
in the market, because markets may be constrained-efficient, that is, they maxin
incomes subject to the information and other barriers that participants face.
example, in rural areas, poverty, low population density, isolated markets, seaso
ity, and highly covariant risk such as widespread crop failures in a given region o
result in high transaction costs, a lack of traditional collateral, variable incomes,
limited opportunities for diversifying risk. These features differentiate rural finan
markets from urban ones and often scare off traditional for-profit financial inter
diaries. They do not, however, entail market failures, because these features resul
high real costs to society that government interventions would also face. At the s
time, there may be clear economic gains that the market has not yet obtained,
example, transactions in which the first entrants bear the full cost of losses but
unable to capture the full benefits of success (box 1).

Careful analyses that identify market failures and specify their causes should
cede appropriate interventions to expand rural incomes. Even if a market failur
identified, direct interventions (through subsidies, credit programs, or institutio
are warranted only if the market failure can be addressed cost effectively; thus,
benefits must exceed the costs. A government failure is not a solution to a ma
failure.

Although policy and regulatory reforms that promote growth are often the m
promising way to reduce rural poverty, special interventions may be required if e
nomic growth is not appropriately shared. These interventions are justified based
social norms rather than on market failures (see figure 2). Interventions in r
financial markets, however, are still warranted only if they are the most cost-effec
means of reducing poverty.

Box 1. *How Imperfect Information Can Generate Market Failures*

It is widely—incorrectly—assumed that if a given activity were profitable, someone in the pri
sector would have done it already. As Besley (1994) notes, "An inefficiency might develop if individ
hang back waiting for others to try things out. The slow diffusion of certain agricultural technolo
has often been attributed to a reluctance to be the first user. An obvious role for governm
intervention is to subsidize early innovators. Thus experiments in institutional design, such as
Grameen Bank in Bangladesh, might serve as prime candidates for subsidization. Such argum
appear only to justify subsidizing new ventures, however, and subsidies should be phased-out al
the way."

Box 2. *Two Perspectives on Market Failure and the Argument for Intervention*

Stiglitz (1993): "There *is* a role for the state in financial markets; it is a role motivated by pervasive market failures. In developing countries, market failures are almost undoubtedly greater than in the more developed countries. . . . While limitations on markets are greater in less developed countries than in developed countries, so too, many would argue, are limitations on government. We have argued that government policies can be designed which are attentive to those limitations. . . . What is clear is that a simple ideological commitment to financial market liberalization cannot be derived either from economic theory or be justified by an examination of a broad base of experience. . . ."

Besley (1994): "In summary, there may be good arguments for intervention, and some may be based on market failure. But as one unpacks each argument, the realization grows that, given the current status of empirical evidence on many relevant questions, it is impossible to be categorical that an intervention in the credit markets is justified. Empirical work that can speak to these issues is the next challenge if the theoretical progress on the operation of rural credit markets is to be matched by progress in the policy sphere."

Government interventions in rural financial markets should aim to remove the causes of market failure or poverty, using the most appropriate mix of *instruments,* such as funding for pilot programs; *institutions,* such as private financial intermediaries, nongovernmental organizations, or state-owned rural financial institutions; and *products,* such as credit, savings, guarantees, and insurance. Where appropriate, market failure caused by imperfect information can be addressed by providing seed capital to establish rural financial institutions in remote areas (box 1).

Interventions should always be designed to complement, facilitate, or improve rural financial markets over the long term (box 2). For example, if an initial cost-benefit analysis suggests that a state-owned rural financial institution is a more cost effective vehicle for promoting rural financial markets than working through private banks or nongovernmental organizations, the government should not later prop up its rural financial institutions with more favorable access to subsidies or concessional funds than are available to other entrepreneurs. On the contrary, competition should be encouraged. Subsidies or grants should generally be restricted to seed capital or be limited by a sunset clause. Finally, the cost of programs to develop rural financial markets should be monitored to assess the cost-effectiveness of the interventions.

There have been widespread failures among state-owned specialized agricultural credit institutions around the world (Adams, Graham, and Von Pischke 1984). These agencies have lacked appropriate governance, capable management, political autonomy, and innovative, efficient operating procedures. They have not addressed information constraints and have been plagued by incentive problems. Taking into account forgone opportunities by private rural financial institutions, the inefficient allocation of resources, and the fiscal costs of propping up loss-making and often

Box 3. *Two Good Reasons for Market Interest Rates: Equity and Efficiency*

Equity: Directed credit programs invariably face the following dilemma: whether to lend to more clients with no subsidy, or to fewer people with a high subsidy per dollar lent. If the issue is perceived as resolving the inadequate *access* to formal credit of the rural masses, then (on equity grounds), the policy should pursue *increased outreach*—a choice that requires eliminating or minimizing the subsidy per dollar lent.

Efficiency: Several studies show that liberalized financial markets generate a more efficient allocation of resources and higher rates of economic growth (King and Levine 1993; Jaramillo, Schiantarelli, and Weiss 1993; McKinnon and Shaw 1976). Other studies point to a positive relationship between savings and real interest rates in developing countries (Fry 1988). The importance of financial institutions in offering and charging positive real interest rates is clearly shown in King and Levine (1993), who find that real growth in gross domestic product during 1974–89 for a sample of 76 countries was more than 2 percent higher for those offering the highest deposit interest rates than for those offering the lowest deposit rates. Indeed, growth was negative for the latter group of countries.

corrupt institutions, the economic costs are likely to have far outweighed the benefits of these public-sector institutions. Even worse, the rich have frequently captured the subsidies, compounding problems of poverty and inequality.

State provision of financial services in rural areas is not recommended unless it meets the following strict governance criteria:

- Fully autonomous management that is held accountable for the bank's financial performance
- Exemption from civil service pay scales to attract and reward quality staff on the basis of the institution's financial performance
- Insulation from staffing pressures by local authorities, for example through autonomous organizational charts with professional qualifications criteria
- The same freedom to set borrowing and lending rates that apply to commercial banks, so that both deposit and lending rates are at market rates, are usually positive in real terms, and provide an adequate spread to cover costs (box 3)
- Application of international best practice prudential regulatory, accounting, and disclosure practices, and therefore the development of a strong management information system by the rural financial institutions, and both off-site supervision and on-site examinations by the same agency that supervises private banks
- A hard budget constraint
- A clear strategy to develop rural financial markets that are supported only with initial, nonexclusive, time-bound, transparently budgeted subsidies.

Measuring the Performance of Rural Financial Intermediaries

Evaluating the effect of a rural credit program on incomes and poverty is very difficult because it is rarely clear what borrowers would have done in the absence of the program. Therefore, practitioners and academics have developed a new framework for assessing the performance of credit programs. This framework rests on outreach and self-sustainability (Yaron 1992a). It argues that rural financial institutions that provide a broad range of services to the targeted clientele in an efficient manner are likely to have the desired impact of expanding incomes and reducing poverty. Therefore, evaluating their performance based on these criteria provides an easily quantifiable proxy of the impact of rural financial intermediation in lieu of a full benefit-cost analysis (figure 3).

Figure 3. *Assessing the Performance of Rural Financial Institutions*

Assessment criteria

Self-sustainability

Subsidy dependence index
Measures subsidies received against interest earned

Examples of subsidies

- Interest rate subsidy on concessionally borrowed funds
- Opportunity cost of equity
- Reserve requirement exemptions
- Free equipment provided by government or donors
- Government's assumption of loan losses
- Free training for staff provided by government or donor
- Government assumption of foreign exchange losses

Outreach to target clients

Outreach index
Evaluates outreach to clients and quality of services offered

Examples of indicators

Market penetration
- Number and annual growth rate of saving and loan accounts
- Value and annual growth rate of the loan portfolio and deposits
- Number of branches and staff

Relative income level
- Value of average loan and range of loan amounts
- Percentage of rural clients
- Percentage of female clients

Quality of services
- Transaction costs to clients
- Flexibility and suitability of services
- Distribution network

Source: Yaron, Benjamin, and Piprek (1997)

Outreach is measured by a hybrid index comprising several indicators, such as the number of clients, the value of the loan portfolio and its annual growth, the percentage of female clients (where social norms discriminate against women), the average loan size (as a proxy for income level of the clientele), and so on.

Self-sustainability is assessed by calculating the subsidy dependence index, that is, the percentage by which the agency's average on-lending interest rate would have to increase to make it self-sustainable (Yaron 1992b; Benjamin 1994). Conventional accounting practices fail to reflect most subsidies received by state-owned rural financial institutions or by nongovernmental organizations and therefore do not show the true social costs of maintaining these intermediaries. Yet without this measure it is impossible to determine whether continuing support for those institutions is warranted. Given the prevalence and extent of subsidies, recognizing the subsidy dependence is essential to evaluating the performance of state-owned institutions and nongovernmental organizations. The subsidy-dependence index is instrumental in:

- Relating the total amount of subsidies received by a rural financial institution to its level of activity, represented by the interest earned on its loan portfolio. This exercise is similar to calculations of effective protection, domestic resource cost or job creation cost. It also captures the notion of matching grants by comparing the value of subsidies received against the income earned from clients in the market place.
- Tracking a rural financial institution's performance in terms of subsidy dependence over time and relative to that of other institutions that provide similar services to a similar clientele.

Successful Rural Financial Institutions

Three Asian rural financial institutions are widely considered successful based on the two primary criteria of outreach and self-sustainability: the Bank for Agriculture and Agricultural Cooperatives in Thailand; the Village Banks, or Unit Desas, of Bank Rakyat Indonesia (BRI-UD); and the Grameen Bank in Bangladesh. All have succeeded in providing financial services at unprecedented levels to millions of rural people. The Grameen Bank has reduced its dependence on subsidies, the Bank for Agriculture and Agricultural Cooperatives has benefited from low to modest subsidies, and the Village Bank program of the Bank Rakyat Indonesia has completely eliminated its dependence on subsidies. These successes contrast sharply with the traditional view that heavy subsidies are inevitable in the provision of financial services to rural entrepreneurs. A variety of mechanisms have enhanced the efficiency of these institutions, including the following:

- A high degree of management autonomy in formulating operational policies

- Policies that provide for staff accountability, investment in human capital, and rewards (monetary incentives and promotions) that are related to sound financial performance and sustainability.
- Innovative, low-cost delivery systems and mobile banking services.
- Innovative and flexible loan terms and conditions adapted to social, economic, and cultural circumstances. For example, all offer weekly or monthly repayment schedules tailored to the clients' cash flow.
- Close monitoring of loan performance; high, on-time collection rates and low loan losses.
- Development of domestic savings accounts to reduce or eliminate the need for donor funds.
- Positive and often relatively high on-lending rates that ensure an adequate spread.
- Control over administrative expenses and effective use of economies of scale.
- Advanced management information systems that facilitate effective planning, control, and timely monitoring of loan repayments.
- Concentration on rural markets that have relatively high population densities.

A close look at the operations of the BRI-UD shows how it has achieved unparalleled success in rural financial intermediation.

Explaining the Success of the Village Banks Program

This successful public entity has succeeded in reaching financial self-sustainability while providing credit and saving services to rural low-income families that had not previously had access to formal financial services. Moreover, it has achieved an unprecedented level of profitability, earning $177 million in profits during 1996, and a return on average assets of more than 5 percent a year between 1994 and 1996. This level of profitability is rarely found even among financial intermediaries that serve clients who borrow and save vastly larger amounts, and who have much more substantial enforceable collateral.

Background

By the mid-1970s Indonesia's directed-credit program aimed at channeling funds to rice farmers (under the BIMAS, or Mass Guidance program), had contributed to making the country self-sufficient in rice production. By the early 1980s, however, the program had become increasingly unsustainable as a result of subsidized interest rates, poor loan repayments, and employee incentives directed toward disbursing credit rather than generating profits. In 1984 the government transformed the op-

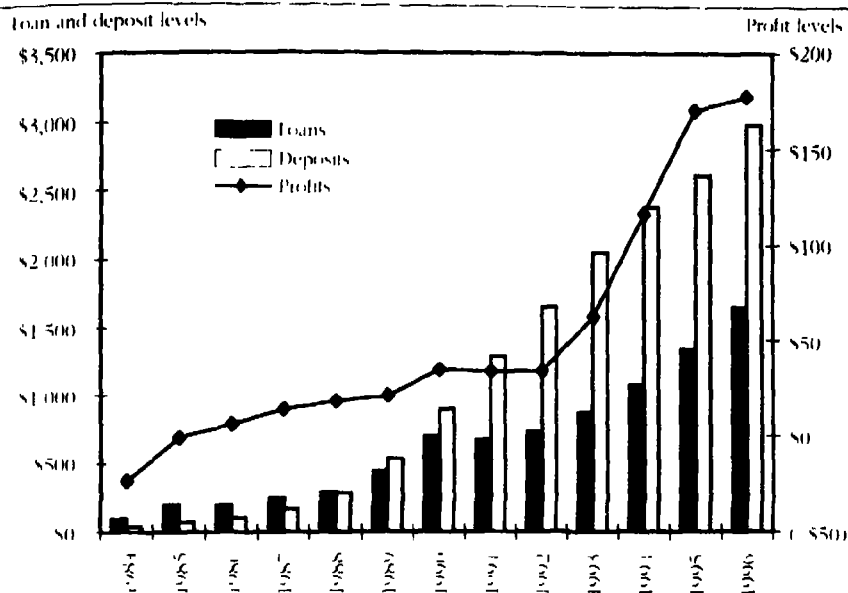
eration into the BRI-UD system and ordered it either to devise a program to provide rural financial services on a self-sustaining basis or to face closure. With a relatively small initial subsidy in 1984, the new BRI-UD became profitable within eighteen months. By 1996 the bank was a global leader in rural financial intermediation, with 2.5 million loan accounts and 16.2 million deposit accounts, combining substantial market penetration among low-income rural clients with sustained profitability (table 1 and figure 4).

The most fundamental policy change was a shift from disbursing credit to motivating loan recovery and mobilizing savings; that is, to genuine rural financial inter-

Table 1. *BRI-UD's Outreach and Financial Self-Sustainability*
(US$)

Outreach	*1985*	*1990*	*199*[illegible]
Average annual loan volume (millions)	162	562	1,17[illegible]
Number of outstanding loans (millions)	1.0	1.9	2.3
Average outstanding loan amount / borrower	162	296	51[illegible]
Average annual deposit volume (millions)	49	685	2,38[illegible]
Number of deposit accounts (millions)	—	7.3	1[illegible]
Average deposit amount per saver		94	16[illegible]
Financial self-sustainability (in percentages unless noted)			
Nominal average yield earned on loan portfolio	27.4	31.5	31.[illegible]
Nominal average interest rate paid on deposits	10.5	11.3	9[illegible]
Nominal interest rate spread	16.8	20.2	21[illegible]
Inflation	4.7	7.4	9[illegible]
Real average yield earned on loan portfolio	21.[illegible]	22.4	20.[illegible]
Real average interest rate paid on deposits	5.6	3.6	[illegible]
Lowest lending rate needed for financial self-sustainability			
Nominal	36.2	27.2	1[illegible]
Real	30.1	18.4	[illegible]
Operating costs as a percentage of			
Average annual net loan portfolio (LP)	20.5	12.9	1[illegible]
Half of the average annual net LP and deposits	31.5	11.6	8[illegible]
Average annual total assets	15.1	8.0	[illegible]
Profits ($ millions)	–0.8	34.3	17[illegible]
Percentage of profitable units	48.3	89.1	9[illegible]
Average ann. deposit volume / average ann. LP volume	0.31	1.22	2.0[illegible]
Subsidy dependence index	32.2	13.7	[illegible]

— Not available.
Source: Yaron, Benjamin, and Piprek (1997).

Figure 4. *BRI-UD Loan and Deposit Growth and Profitability (US$ millions)*

Source: Bank Rakyat Indonesia's financial statements and authors' calculations.

mediation. To broaden and diversify its clientele, the bank targeted the low-income rural population and offered loans for all income-generating activities—a sharp departure from the traditional pattern of lending solely to the agricultural sector, which has accounted for a shrinking share of gross domestic product. BRI-UD underpinned these policy shifts with substantially higher loan and deposit interest rates, while maintaining a sufficient interest-rate spread to cover the high costs of servicing small loans and deposits. The average annual yield on loans has been about 32 percent in recent years; average annual financial costs have been about 10 percent. Innovative incentive systems were introduced for both clients and employees to encourage timely and complete loan repayment (Chaves and González-Vega 1996).

An Autonomous Organizational Structure

The Village Bank system functions as an independent profit center within the Bank Rakyat Indonesia. The Village Banks are free to set their own loan terms, although transfer prices (discussed below) are negotiated with the bank. BRI-UD has developed its own management tools, including an efficient management information system

to assess performance and a sophisticated employee incentive system to encourage profitability, loan recovery, and savings mobilization.

Since 1984, 10 percent of each unit's annual profit has been distributed to employees as a reward for achieving good collection rates. Bonuses are paid early in the year and are capped at 1.5 month's salary per employee. Because about 96 percent of the units were profitable in 1996, a similar percentage of more than 21,000 employees and trainees are benefiting from this program as well as from additional bonuses for achieving goals that are earned in routine competitions between units. The importance of these incentive schemes cannot be overstated; they clearly set BRI-UD apart from government development banks elsewhere that remunerate their employees on the basis of inadequate civil service pay scales.

Innovations for Rural Customers

The loan application process takes about a week for a new borrower and less time for a repeat customer. Loans are extended on an individual basis, and generally have a maturity of 18 months, with monthly repayments. Collateral is desirable but not mandatory. Loan delivery systems incorporate cost-minimizing features. For example, paperwork is kept to a minimum, and where the volume of business is relatively small, mobile offices provide limited services to clients in outlying areas once or twice a week.

The small average size of loans and relatively high cost of legal procedures make foreclosure prohibitively expensive (although warranted in certain cases to achieve a demonstration effect). Thus the focus is on quality at entry and appropriate incentives for repayment:

- Applicants are prescreened based on available information (gathered from peer and from village leaders), and proposed investments are evaluated
- Clients are given a substantial incentive to repay through both interest rebates (of about 12 percent a year) and access to additional larger loans contingent on timely repayment.
- Staff incentives are linked to the performance of the loan portfolio, so clients are monitored more closely than is usual.
- Loans are priced to encourage more selective choices of investments and to promote credibility in the institution.

Because BRI-UD is not viewed as just another transient government program, borrowers have a greater incentive to repay their loans and depositors place greater trust and confidence in the institution. Fieldwork throughout the country since 1982 has pointed to extensive demand in rural areas for reliable financial savings facilities especially for liquid savings accounts (Robinson 1994). Four savings instruments with interest rates that varied substantially with account size and liquidity were of-

fered in 1986 as part of the new rural savings program. This program was the flagship of the bank's revamped effort to provide services to rural clients.

Managing Banking Operations

The BRI-UD's loan-loss treatment is very conservative compared with most state-owned rural financial institutions in other countries. In addition to general loan loss reserves of 3 percent against all outstanding loans that are not yet due and payable, there are reserves of 50 percent against loans less than three months overdue, and reserves of 100 percent against loans that are three months to a year overdue. Loans that are more than one year overdue are fully written off. Asset classification is also conducted conservatively with a view to avoiding hidden rescheduling, or "evergreening," of the portfolio.

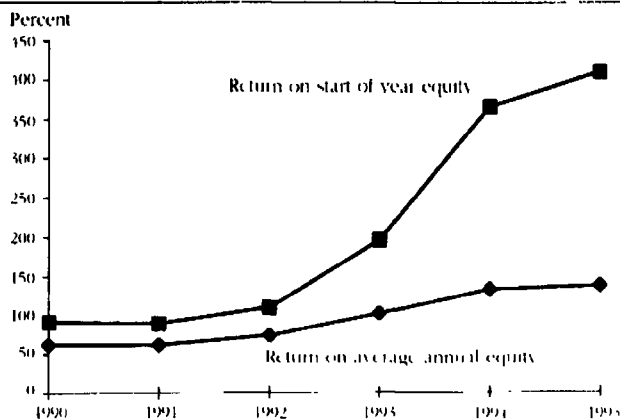
One of the principal advantages of belonging to a nationwide branch network is that the Bank Rakyat Indonesia system serves as a clearing house between cash-surplus and cash-deficit units. Fund transfers carry an interest rate—the transfer price—which is adjusted periodically according to the bank's overall liquidity position. The transfer price is usually set slightly higher than the top savings rate offered at the units so that those with a surplus of funds can at least cover their interest costs and are not discouraged from mobilizing savings (Charitonenko, Patten, and Yaron 1998).

To improve the quality of bookkeeping and customer service, Bank Rakyat Indonesia has developed a stand-alone, personal computer-based system. By the end of 1995 computerization had been introduced in about 89 percent of the 3,135 Village Banks, facilitating the units' bookkeeping and management practices.

Measuring Financial Performance

The two most widely used financial ratios for measuring a financial institution's performance are its return on equity and return on assets. In addition, the subsidy dependence index is increasingly being used to evaluate state-owned financial institutions and nongovernmental organizations that provide financial intermediation services because the profitability of these intermediaries often depends on their access to subsidies.

In the case of BRI-UD, a further refinement is required. Typically, the return on equity compares an institution's net income in a given fiscal year with its average equity during that fiscal year. In the absence of more detailed data (on the timing and amount of capital injections and dividend payments), average annual equity is generally calculated simply as the sum of start-of-year equity plus year-end equity, divided by two. Complete information is available for the BRI-UD, however: there have been no capital injections or grants; and every year on January 1, all profits earned during the preceding year are transferred to Bank Rakyat Indonesia's general

Figure 5. *BRI-UD's Return on Equity, Measured against Average Annual Equity and Start-of-Year Equity*

Source: Bank Rakyat Indonesia's financial statements and authors' calculations.

account. Because the units do not retain their earnings, it is more meaningful to calculate the return based on start-of-year equity. This refinement has major implications for measuring profitability, as shown in figure 5. The system has been exceptionally profitable by any banking standards. Whereas banks in low-inflation countries might earn 15 to 20 percent (after tax) on their average annual equity, BRI-UD earned more than 60 percent on its average equity in 1990 and 1991. By 1995 the figure had more than doubled to 136 percent. When net income is measured against start-of-year equity, the return on equity increases to about 90 percent in 1990 and 1991, rising to an astounding 407 percent by 1995 (a net income of Rp. 403 billion on a start-of-year capital of Rp. 99 billion).

The relatively low ratio of equity to assets is an important factor in explaining these high returns on equity. If BRI-UD were to become an independent microfinance intermediary, rather than a profit center within a larger bank, it would have to maintain a significantly higher equity-to-assets ratio (say, 15 percent instead of only 1 percent in 1995). The return on assets thus offers a more meaningful indicator of performance.

In competitive financial systems, a 1 percent return on assets is considered an indication of sound financial performance; figures of 2 to 3 percent are often recorded in the better performing commercial banking systems in emerging markets. By contrast, BRI-UD's pre-tax return on assets reached 6.1 percent in 1995, more than double the 2.6 percent return earned in 1991.[1] Thus, whatever common financial indicator is used to assess their performance, the units have earned returns on rural financial intermediation that are well above those in the banking industry

Several rural financial institutions in developing countries have reported adequate financial returns using the returns on equity and returns on assets, even though they are in fact dependent on subsidies (these include the Agricultural Development Bank of Jamaica, the Caisse Nationale de Credit Agricole in Morocco, the Grameen Bank in Bangladesh, and the Bank for Agriculture and Agricultural Cooperatives in Thailand). Because the BRI-UD was built on an earlier program's infrastructure of rural branches, it was able to shed its subsidies in only three years (figure 6). In 1995 it achieved a negative subsidy dependence index; the units could have reduced the yield on their loan portfolios by 44.5 percent (from 31.6 to 17.5 percent) and still remained independent of subsidies and earned an adequate market rate of return on equity.[2] This indicates the "real" profitability that has resulted from effective rural finance intermediation and underscores the tremendous potential for efficient and profitable rural finance in other countries.

Although more in-depth impact evaluations are still required, borrowers enjoyed 25 percent growth in real profits, 21 percent growth in household income, and 18 percent growth in employment per enterprise, according to earlier studies (Sutoro and Haryanto 1990, as cited in Boomgard and Angell 1994; Patten and Rosengard 1991). Moreover, women have had greater access to banking service (25 percent) than is common in the Indonesian banking system (Reed and Befus 1993). The profits have also attracted competition from both private and public lenders, particularly at the higher end of the rural financial market, prompting BRI-UD to cut its on-lending rate on its larger loans by 7 percentage points; in some areas it also faces competition for deposits (Ravicz 1998).

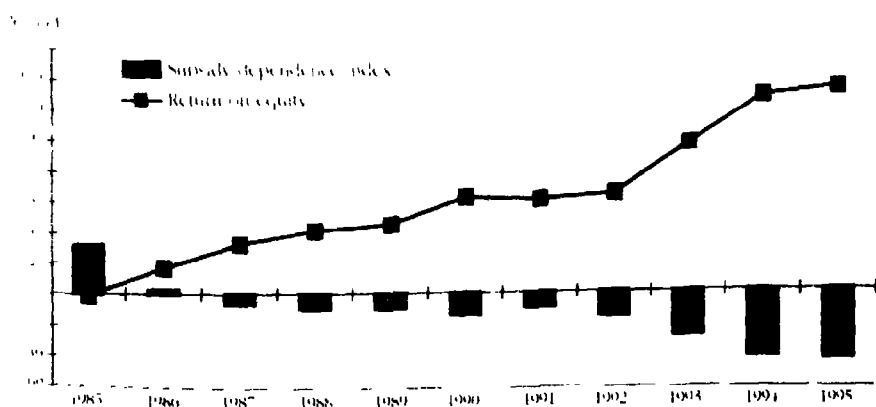
Figure 6. *BRI-UD's Subsidy Dependence Index and Return on Average Equity*

Source: Bank Rakyat Indonesia's financial statements; authors' calculations.

Table 2. *Factors That Explain the Demise of BIMAS and the Success of BRI-UD*

Attribute	*BIMAS Credit Program, 1970–84*	*BRI-UD, 1984–present*
Institutional objective	Disbursement conduit for subsidized credit	Profitmaking, full-service rural bank
Financial autonomy	BIMAS windows in BRI branches, with accounts subsumed in BRI branches' financial statements	Distinct profit or loss centers with separate financial accounting
Operational autonomy	Limited—borrowers chosen in practice by extension workers of the Ministry of Agriculture, which certified BIMAS participants	Full—borrowers selected on the basis of the financial viability of their farm or off-farm enterprise
Staff evaluation and accountability	Primarily based on the volume of disbursements or on hectares covered	Primarily based on the profitability of Unit Desas
Staff incentives	Civil service-like flat salary structure, promotions	Profit-related bonus incentives, promotions
Target market	Rice farmers	Any income-generating enterprises
Client incentives	Timely payment incentive effectively none. Penalty for delinquency: curtailment of further loans, although not well enforced	Timely payment incentive: substantial interest rebate, larger follow-on loans. Penalty for delinquency: curtailment of further loans; incentives well monitored and enforced
Interest rates	12% (subsidized), below both the inflation rate and the interest rate paid on small savings deposits	Around 30% (not subsidized), well above both the inflation rate and the interest rates paid on small savings deposits
Main sources of funds	Concessional lines of credit, plus grants	Client deposits at market rate of interest
Dealing with losses	Soft budget constraint: operating losses covered by government	Hard budget constraint: loss-making operations suspended
The bottom line	Heavy losses and subsidy dependence	Large profits and self-sustainability

Source. Authors' findings.

Several insights into replicable findings can be gathered by contrasting the Village Bank system with its predecessor, the BIMAS credit program, which incorporated most features of traditional credit programs in other countries (table 2). That is, it offered targeted credit at below-market rates of interest and focused primarily on the volume of disbursements rather than on loan recovery and institutional viability. The lack of attention to the program's long-term institutional viability encouraged adverse incentives on the part of staff and clients that ultimately led to its demise. Considerable attention must be paid to creating a conducive institutional framework for a public intervention to succeed and therefore for the market to develop. The key elements of such an environment include a hard budget constraint; full operational autonomy (that is, insulation from political interventions); skillful management of information; and a careful alignment of staff and client incentives with long-term institutional objectives.

What Has Happened to the Profits?

The answer to this question is far from reassuring. The vast profits have been used to cross-subsidize Bank Rakyat Indonesia's wealthier clients. In fact, even as the Village Bank system succeeded, the rest of the bank continued to suffer from low recovery rates. This issue is of the utmost importance, because the enormous size of the cross-subsidy results in regressive income redistribution; year after year, small-scale entrepreneurs subsidize their more affluent countrymen. The rural lending scheme's very success may have reduced the pressure on the parent bank to achieve an equivalent level of efficiency. In 1984, when BRI-UD was handed an ultimatum to become self-sustaining or face closure, it may have been expedient to leave the bank's traditional lending to influential borrowers essentially unchanged. But it is clearly time to review these arrangements in light of their substantial economic costs to the country and their perverse effect on poverty reduction objectives.

Considerations for the Future

It is premature to assess the impact of the current financial crisis in Indonesia on the BRI-UD and its clients. Clients have not borrowed in foreign-denominated loans; the portfolios of the individual units were of excellent quality and high liquidity before the crisis —with a 55 percent loan to deposit ratio in 1996—and BRI-UD has enjoyed a flight to quality on the deposit side. Profits had not declined significantly as of 1997, largely due to high transfer prices late in the year for units with surplus liquidity. As an institution, however, Bank Rakyat Indonesia faces rising arrears and foreign exchange losses, reinforcing arguments in favor of broader reforms.

Nonetheless, the BRI-UD experience demonstrates not only that financial servic can be extended to low-income rural clients at lower costs than previously thoug possible, but that they can in many cases be provided while significantly reducing even eliminating the need for any subsidies. The challenges that remain are those strengthening the policy environment, improving the legal and regulatory fram work, and adopting appropriate governance arrangements, management principl and operating procedures for interventions that reflect a new and more promisi approach to rural finance.

The findings related to the level of subsidy independence are extremely importa indications of the potential for improvement in the operations of numerous speci ized agricultural credit institutions that still have little outreach to the target cliente and are heavily subsidized. Government policymakers, managers of state-owned r ral financial institutions, and directors and managers of for-profit financial interm diaries generally believe that rural financial intermediation is not a profitable prop sition and that subsidies are essential to compensate for mandated thin spreads an large loan losses. The experience of the Village Banks program shows that losses a not inevitable and that substantial and consistently increasing outreach can be achiev in rural financial markets in a self-sustaining manner. Indeed, the key lesson is cr tal clear: rural finance can be highly profitable, even when it serves low-income client

Note

Jacob Yaron is rural finance adviser in the Private Sector Development and Rural Developme Departments of the World Bank. McDonald Benjamin is a financial economist in the East Asia Pacific Region Financial Sector Development Unit of the World Bank. Stephanie Charitonenko a consultant in the Private Sector Development Department at the World Bank and a gradu student in agricultural economics at Michigan State University. This article draws on materi Yaron, Benjamin, and Piprek (1997)

1. Although BRI-UD, as a profit center within BRI, is not subject to corporate taxation, if a tax of 30 percent were applied, its return on investment would decline to 1.8 percent and [illegible] percent for 1991 and 1995, respectively.

2. A negative subsidy dependence index of 44.5 percent has no equal in rural or microfinanc

References

The word "processed" describes informally reproduced works that may not be commonly avail through library systems.

Adams, Dale W., Douglas Graham, and J.D. Von Pischke, eds. 1984. *Undermining Rural Dev ment with Cheap Credit*. Boulder, Colo: Westview Press.

Benjamin, McDonald Jr. 1994. *Credit Schemes for Microenterprises. Motivation, Design and Via ity*. Ph.D. dissertation, Georgetown University, Washington D.C.

Berenbach, Shari, and Craig Churchill. 1997. *Regulation and Supervision of Microfinance Institutions: Experience from Latin America, Asia, and Africa.* Occasional Paper 1. The MicroFinance Network, Washington, D.C.

Besley, Timothy. 1994. "How Do Market Failures Justify Interventions in Rural Credit Markets?" *The World Bank Research Observer* 9(1, January):27–48.

Boomgard, James, and Kenneth Angell. 1994. "Bank Rakyat Indonesia's Unit Desa System: Achievements and Replicability." In Otero and Rhyne.

Charitonenko, Stephanie, Richard Patten, and Jacob Yaron. 1998. "Bank Rakyat Indonesia—Unit Desa 1970–1996." Draft Case Study for the Sustainable Banking with the Poor Project. World Bank, Africa Technical Department, Washington, D.C. Processed.

Chaves, Rodrigo, and Claudio González Vega. 1996. "The Design of Successful Rural Financial Intermediaries: Evidence from Indonesia." *World Development* 24:65–78.

David, Cristina, and Richard Meyer. 1984. "Measuring the Farm Level Impact of Agricultural Loans." In Von Pischke, Adams, and Donald.

Fleisig, Heywood, and Nuria de la Pena. 1996. "Creating a Legal and Regulatory Framework to Promote Access to Rural Credit." Center for the Economic Analysis of Law, Washington, D.C. Processed.

Fry, Maxwell. 1988. *Money, Interest, and Banking in Economic Development.* Baltimore, Md.: Johns Hopkins University Press.

González Vega, Claudio. 1984. "Credit Rationing Behavior of Agricultural Lenders: The Iron Law of Interest Rate Restrictions." In Adams, Graham, and Von Pischke.

Jaramillo, Fidel, Fabio Schiantarelli, and Andrew Weiss. 1993. "The Effect of Financial Liberalization on the Allocation of Credit: Panel Data Evidence for Ecuador." Policy Research Working Paper 1092. World Bank, Country Economics Department, Washington, D.C. Processed.

Khan, Azizur Rahman. 1977. *Poverty and Landlessness in Rural Asia.* Geneva: International Labour Organization.

King, Robert, and Ross Levine. 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics* 108 [illegible]

Ladman, Jerry, and Ronald Tinnermeier. 1984. "The Political Economy of Agricultural Credit: The Case of Bolivia." In Adams, Graham, and Von Pischke.

McKinnon, Ronald, and Edward Shaw. 1976. *Money and Finance in Economic Growth and Development: Essays in Honor of Edward Shaw. Proceedings of the Conference held at Stanford University.* New York: M. Dekker.

Otero, Maria, and Elisabeth Rhyne, eds. 1994. *The New World of Microenterprise Finance: Building Healthy Financial Institutions for the Poor.* West Hartford, Conn.: Kumarian Press.

Patten, Richard, and Jay Rosengard. 1991. *Progress with Profits: The Development of Rural Banking in Indonesia.* San Francisco: ICS Press.

Ravicz, Marisol. 1998. "Searching for Sustainable Microfinance: A Review of Five Indonesian Initiatives." Policy Research Working Paper 1878. World Bank, Development Research Group, Washington, D.C. Processed.

Reed, Larry, and David Befus. 1993. *Transformation Lending: Helping Microenterprises Become Small Businesses.* GEMINI Working Paper 37. Washington, D.C.: United States Agency for International Development.

Robinson, Marguerite. 1994. "Savings Mobilization and Microenterprises Finance: The Indonesian Experience." In Otero and Rhyne.

Schiff, Maurice, and Alberto Valdés. 1992. *The Plundering of Agriculture in Developing Countries.* Washington, D.C.: World Bank.

Stiglitz, Joseph. 1993. "The Role of the State in Financial Markets." Paper presented at th Annual World Bank Conference on Development Economics, Washington, D.C., May 3–4 Processed.

———. 1996. "The Role of Government in Economic Development." Paper presented at th Annual World Bank Conference on Development Economics, Washington, D.C., April 25–2(Processed.

Sutoro, Ann Dunham, and Roes Haryanto. 1990. *KUPEDES Development Impact Survey.* BRI Brief ing Booklet. Jakarta: Bank Rakyat Indonesia.

Virmani, Arvind. 1982. "The Nature of Credit Markets in Developing Countries: A Framework fc Policy Analysis." Staff Working Paper 524. World Bank, Public and Private Finance Divisior Development Economics Department, Washington, D.C. Processed.

Vogel, Robert. 1984. "Savings Mobilization: The Forgotten Half of Rural Finance." In Adam Graham, and Von Pischke.

Von Pischke, J. D., Dale Adams, and Gordon Donald, eds. 1983. *Rural Financial Markets in Deve oping Countries.* Baltimore, Md.: Johns Hopkins University Press

World Bank. 1993. *A Review of Bank Lending for Agricultural Credit and Rural Finance (1948–1992 World).* Report 12143. Washington, D.C.

Yaron, Jacob. 1992a. *Successful Rural Finance Institutions* World Bank Discussion Paper 150. Washington, D.C.

———. 1992b. *Assessing Development Finance Institutions A Public Interest Analysis.* World Bank Discussion Paper 174. Washington, D.C.

Yaron, Jacob, McDonald Benjamin, and Gerda Piprek. 1997. *Rural Finance Issues, Design, and Best Practices.* Environmentally and Socially Sustainable Development Studies and Monographs Series 14. Washington, D.C.: World Bank.

The Role of Long-Term Finance: Theory and Evidence

Gerard Caprio, Jr. • Asli Demirgüç-Kunt

Improving the supply of long-term credit to industrial firms is considered a priority for growth in developing countries. A World Bank multicountry study looks at whether a long-term credit shortage exists and, if so, whether it has had an impact on investment, productivity, and growth. The study finds that even after controlling for the characteristics of individual firms, businesses in developing countries use significantly less long-term debt than their counterparts in industrial countries. Researchers are able to explain the difference in debt composition between industrial and developing countries by firm characteristics; by macroeconomic factors; and, most importantly, by financial development, government subsidies, and legal and institutional factors.

The analysis concludes that long-term finance tends to be associated with higher productivity. An active stock market and an ability to enter into long-term contracts also allow firms to grow at faster rates than they could attain by relying on internal sources of funds and short-term credit alone. Importantly, although government-subsidized credit markets have increased the long-term indebtedness of firms, there is no evidence that these subsidies are associated with the ability of firms to grow faster. Indeed, in some cases subsidies are associated with lower productivity.

The popular view holds that financial markets in developing economies are highly imperfect and, in particular, that the alleged scarcity of long-term finance is a key impediment to greater investment and growth. Indeed, a significant part of the lending by the World Bank and other multilateral development banks is aimed at correcting for the dearth of long-term credit through the creation and encouragement of development finance institutions (DFIs) that could lend funds through loans from financial intermediaries and commercial banks, and recently through guarantees that lengthen the maturity of loans. Yet a recent strand of the finance literature has been studying the forces that determine the maturity structure of a firm's debt (Berglof and von Thadden 1993; Diamond 1991, 1993; Rajan 1992). In those models

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 171–89

long-term debt is not necessary for acquiring physical capital and may depress a firm's performance. These analysts would view policy-induced changes in the term structure of finance generally—if not uniformly—with great skepticism.

Notwithstanding the difference of views, attempts to cure the alleged scarcity of long-term credit in developing countries have been plentiful and expensive. By the early 1980s many DFIs were experiencing significant portfolio problems. A 1974 World Bank study of delinquency rates in agricultural lending institutions reported that the average arrears rate was 41 percent. A 1983 report indicated that 39 percent of all DFIs had serious portfolio problems, while another 53 percent faced moderate problems, many of which became more severe in the late 1980s and resulted in a wave of failures by DFIs (Siraj 1983).

Furthermore, in many cases, long-term directed credit failed to reach the intended beneficiaries (Atiyas 1991; World Bank 1989). Once these directed credit programs were established, governments found it politically difficult to reduce support for them, regardless of their cost and inefficiency. Prompted by these problems, the World Bank adopted new guidelines governing lending to DFIs, and World Bank loans to these institutions dropped dramatically, from 11 percent of new credits in fiscal 1989 to only 2.4 percent in fiscal 1993. Controversy continues, however, and both the World Bank and the development community at large are reevaluating mechanisms aimed at increasing the availability of long-term finance or lessening the constraints imposed by its absence.

Although aggregate data and anecdotal evidence suggest that there is less long-term credit in developing countries, even in those countries with low or moderate inflation rates, until recently no attempts had been made to examine the evidence more systematically to see if the scarcity merely reflected the characteristics of firms in poorer economies. This lacuna was understandable even five years ago, because the data were not available. Recently, however, this gap has been filled, first by the availability of data in emerging markets for the top tier of firms listed on stock exchanges and second by various surveys of listed (and unlisted) firms in selected countries whose governments have sought to understand the impact of a variety of policies. Armed with this data, a number of studies have appeared in the last year, and answers are now available to fill an important gap in our knowledge.

Financing Decisions and Debt Maturity

What does it mean to say that long-term credit is scarce? Typically this question has been answered by asking firms to identify the important constraints on their operations; credit—usually long-term credit—regularly is at or near the top of the list. This approach is unsatisfactory, however, not least because it is unclear what the respondents imagine they will pay for credit. Moreover, it is unclear under what type

of financial system they would be able to obtain short- or long-term credit. Even the most advanced financial system will find some borrowers uncreditworthy or will lend them much less than they might desire or at higher interest rates than they would like. In the case of riskier firms, loans at average market rates are attractive precisely because they convey a subsidy in the form of a lower risk premium than the market would grant them. Whenever the lack of long-term credit constrains many firms from expanding, there are three potential sources of the credit constraints: first, macroeconomic factors that limit supply; second, institutional factors that are specific to the financial sector (often dubbed market imperfections); and third, the characteristics of the firms, or classes of firms, in the country.

One way to interpret scarcity then is by relative access to credit. That is, there is scarcity to the extent that developing country firms find it more difficult to borrow money than do similar firms in industrial countries. In this relative sense, if there is a scarcity, there may be a potential correction. To be sure, any correction may be difficult. For example, it is argued (and confirmed below) that a leading cause for the absence of long-term finance is high inflation and unstable macroeconomic policies. Attempts to increase the supply of long-term credit without addressing the inflation problem could easily prove to be short-lived or costly. Similarly, high real interest rates may reduce the effective demand for credit. Entrepreneurs will say they want more credit, but not at the market price. If the yield curve is upward sloping (meaning that long-term interest rates are higher than short-term rates), the demand for long-term credit will tend to suffer most. Again, addressing the factors that account for high real interest rates may in the long run be more useful than attempts to force banks to make longer-term commitments (Brock 1995). In the 1980s Chile succeeded both in tackling the factors underlying high real rates (an overvalued exchange rate and insolvent banks) and, by moving to a fully funded pension system, in creating a natural source of long-term finance without interfering in credit and investment decisions. The studies reviewed here finesse this issue by limiting the study to countries with relatively stable macroeconomic environments.

Institutional factors generally affect borrowers only until funds are disbursed but are crucial during all phases of a credit relationship for providers of funds, who are concerned about the return on their investment. Diamond (1991) points out that banks use short-term credit as a way to control borrowers and that they tend to use this instrument more frequently in cases in which the financial infrastructure is underdeveloped. Thus if information systems or contract enforcement mechanisms are absent, or accounting and auditing techniques are not adequate, lenders will be unwilling to enter into long-term loans. Ignoring this deficiency and establishing government banks for long-term credit is faster, cheaper, and less difficult than trying to address the information or contract enforcement problems, but government banks will have to cope with the same issues and may have additional incentive problems as well.

Finally, the maturity structure differs within an economy depending on the characteristics of domestic firms. Below we review the importance of these firm-specific factors that affect access to long-term finance.

Access within Countries: The Relevance of Firm Characteristics

In the aftermath of the seminal Modigliani-Miller (1958) article, which found that the value of a firm was not affected by its mix of financing, the study of financing choices initially received little attention. As economists and finance experts have renounced the simplifying assumptions of this classic framework, however, they have developed a literature on the maturity structure of firm financing, stressing the different roles played by long- and short-term finance. This literature emphasizes that short-term debt has three effects: it permits loans to be repriced to reflect new information; it increases efficiency by allowing uneconomic projects to be terminated; and it gives managers and owners strong incentives to avoid bad outcomes. In contrast, long-term debt protects the firm from liquidation by imperfectly informed creditors and prevents opportunistic creditors from using the threat of liquidation to expropriate the profits of healthy firms.

Several factors determine the optimal mix of long- and short-term debt. These include the firm's credit rating, its portfolio of growth opportunities, the profitability of the project, the ability to fund the project through retained earnings, the liquidation value of the assets, the perceived accuracy of financial information, the firm's size and age, and the level of banking competition. Valuable assets that can serve as collateral ease borrowing constraints considerably. According to Myers (1977) firms can also use their "growth opportunities" as collateral. Firms whose principal asset is the present value of growth opportunities may not be able to borrow that easily however, because the owner-managers have greater opportunities to divert resources for their own use. As Myers notes, the firm can limit this problem by carrying less debt, by including restrictive covenants in its debt contracts, or by borrowing more short-term debt (which permits the creditors to detect opportunistic behavior relatively quickly). In developing countries, one might expect to find more firms relying on growth opportunities, so this diversion problem could be significant. Moreover since it is difficult in lower-income countries both to sell shares of stock (one way to lower debt-equity ratios) and to enforce contracts (because regulatory mechanisms are typically less developed), businesses can be expected to use more short-term debt By sequencing a series of short-term loans, bankers retain control over their clients because the option to discontinue rolling over these loans is easier to exercise and is a more credible near-term threat.

Hart and Moore (1995) find that the faster the returns to investment are realized, the shorter the optimal payment structure will be. Empirically, this suggests a particular relationship among the maturity of debt, purpose of the loan, and the nature

of the firm's assets. Long-term loans are usually used to acquire fixed assets, equipment, and the like. Short-term loans, on the other hand, tend to be used for working capital, such as payroll, inventory, and seasonal imbalances. Collateral usually consists of such things as inventories or accounts receivable. In other words, firms will tend to match the maturity of their assets and liabilities; only firms with long-term assets will tend to have a longer debt maturity structure. If this tendency is born out in developing country experience, it suggests that the most effective way to deal with the market allocation of credit is to take account of the structure of the firms' assets. A program to extend long-term credit to firms with short-term assets may not be welcomed, as it is inconsistent with the desire to balance the maturity of assets and liabilities.

The size of the firm is another key variable. Indeed, the desire to get more credit—particularly long-term credit—to small firms is a justification for a number of credit market interventions. In general there tends to be less information about small firms, not only because they are new, but also because such information is costly to obtain. Thus, even in the most developed financial systems, small- and medium-size enterprises tend to get a larger part of their external financing from banks. Banks overcome some information problems by developing long-term relationships with smaller firms.

The point is that firms in developing countries may have less long-term debt than firms in developed countries simply because they have different characteristics rather than because of deficiencies in credit markets. Moreover, comparisons of debt maturity structures in different countries are more likely to be informative if researchers control for these parameters.

Although numerous empirical papers test the implications of capital structure models, attention has turned only recently to empirical determinants of debt maturity (see Harris and Raviv 1990 for a review of the literature). Titman and Wessels (1988) find that highly leveraged firms tend to issue more long-term and short-term debt but that the mix varies according to the firm's characteristics. Barclay and Smith (1995) report that large firms as well as those that have few growth options have more long-term debt, a finding confirmed by Stohs and Mauer (1996), who note that larger, less risky firms with longer-term asset maturities use longer-term debt.

In these studies, based on U.S. data, the link that stands up most clearly is the matching of firm assets and liabilities. This finding is quite robust in Italy and the United Kingdom (Schiantarelli and Srivastava 1996), where it is also clear that firms with higher profits have access to more long-term credit. World Bank research using developing country data generally confirmed these results. Maturity matching also is evident in Colombia (Calomiris, Halouva, and Ospina 1996), India (Schiantarelli and Sembenelli 1996), and Ecuador (Jaramillo and Schiantarelli 1996). If maturity matching represents a tendency in both industrial and developing country markets, attempts to stimulate long-term finance may prove to be excessive; firms may take

on long-term debt only if it fits their balance sheet structure, and perhaps only if long-term debt is subsidized, meaning that they can take advantage of a lower risk premium than is available from the market. These country studies confirmed that where financial markets are free from government intervention, they provide more long-term finance to better quality firms, and attempt to monitor lower quality firms more closely by using short-term debt.

The government's decision to intervene in credit markets should depend on the link between long-term credit and the firm's performance as well as on equity considerations. Lack of collateral as well as the age of the business may be factors where small firms find it difficult to obtain long-term credit, as in Ecuador, where only 11 percent of very small firms and 17 percent of small firms reported long-term debt every year (1984–88). Larger firms in Ecuador tended to be more profitable, suggesting that the allocation of credit favored firms with better balance sheets. The allocation could also reflect the economic and political power of such firms. A more disturbing aspect was that, regardless of the size of the firm, the amount of long-term credit obtained was unrelated to past profits. Whether this reflects a market failure, the limits of banking (bankers can pick the class or industry, but not individual winners and losers), or excessive intervention (a substantial portion of the debt was subsidized) is not clear.

Access across Countries: The Relevance of Institutional Factors

Financial theory suggests that a major factor in the choice of capital structure is the cost of contracting between firms and their providers of capital. It is the institutions in the economy—legal or financial—that facilitate monitoring and enforcement of financial contracts.

For example, when the legal system is costly or inefficient, short-term debt is more attractive than long-term debt (Hart and Moore 1995; Bolton and Scharfstein 1993). Diamond (1991, 1993) also emphasizes the importance of clear legal rules to govern contract enforcement. This implies that if complicated loan covenants (to anticipate a variety of future outcomes) could be enforced at a lower cost, the risk for the lender would be reduced and the willingness to expand the supply of long-term debt would increase.

FINANCIAL INSTITUTIONS. Two types of institutions—financial intermediaries and stock markets—directly influence an enterprise's choice of financial structure. Financial intermediaries have a comparative advantage in monitoring borrowers because, as Diamond (1984) argues, bankers have economies of scale in obtaining information. They may also have greater incentives to use the information to discipline borrowers than do small investors. By collecting information, monitoring borrowers, and exerting corporate control, a developed banking sector can facilitate acc[illegible]

to external finance—especially long-term finance—by smaller firms that have limited access to alternative means of financing due to information costs.

Large stock markets allow entrepreneurs the opportunity to substitute equity for long-term debt. Moreover, the prices quoted in financial markets also transmit information that is useful to creditors (Grossman 1976; Grossman and Stiglitz 1980). This revelation may make lending to a publicly quoted firm less risky and thereby increase its ability to obtain long-term credit.

GOVERNMENT PROVISION. In an effort to promote the availability and use of long-term debt, governments may adopt policies that direct or subsidize long-term capital to favored firms or sectors. Directed credit policies include preferential discount lines from the central bank, portfolio restrictions on private commercial banks, guaranteed credit for public enterprises, and credit lines through development banks. These programs need not always involve financial subsidies, but they frequently do. The degree of these distortions varies from country to country.

For example, Atiyas (1991) and World Bank (1989) provide evidence that directed credit often fails to reach its intended beneficiaries. In many cases, such programs are used not to correct the inadequacies of financial markets, but to channel funds to priority sectors whether or not these are the most productive investments. In many countries, including Brazil, Colombia, India, Kenya, Mexico, and Turkey, government interventions have generated large costs by funding inefficient borrowers and crowding out private credit intermediaries.

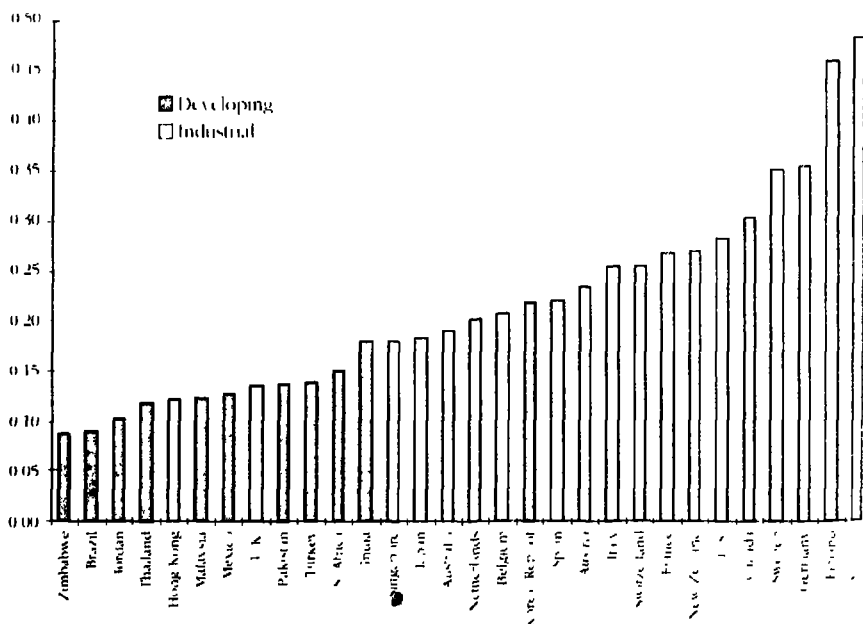
Directed credit programs did achieve their legitimate objectives in a few cases. In successful interventions, as in Japan, Korea and Singapore, credit policy priorities are determined as part of a national plan with broad participation. Commercial standards are applied within the priority sectors; once priorities have been established, lending decisions by agencies are shielded from public pressures; interventions that do not work are discontinued (Calomiris and Himmelberg 1993; Stiglitz and Uy 1996). Where political systems do not allow government authorities to develop and implement effective plans for the distribution of industrial credit, however, success may be difficult to achieve.

THE INSTITUTIONAL ENVIRONMENT. Several studies explore the effect of the institutional environment on the choice of debt. Hoshi, Kashyap, and Scharfstein (1990) show that membership in industrial groups linked to banks reduces financial constraints on Japanese firms, and Schiantarelli and Sembenelli (1996) find the same benefits flow to Italian firms that are members of large national groups. Calomiris (1993) examined the effect of differences between the banking systems of the United States and Germany and argued that regulatory limitations on the scale and scope of U.S. banks hampered financial coordination and increased the cost of capital for industrialization. Rajan and Zingales (1995) and Demirgüç-Kunt and Maksimovic

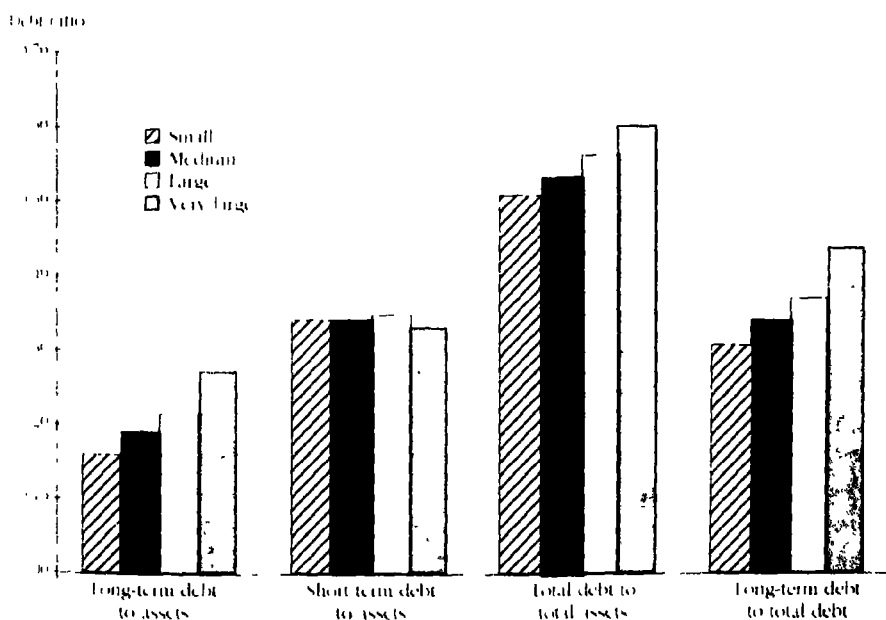
(1994) compare the capital structure of firms in five industrial countries and t developing countries, respectively, and find that institutional differences were cr cial in understanding the determinants of capital structure.

Because of data constraints, however, systematic cross-country empirical stud have been few and recent. Demirgüç-Kunt and Maksimovic (1996a, 1996b), wl look at debt-equity ratios in 30 industrial and developing countries from 1980 1991, find that access to an active stock market increases firms' ability to borro especially in countries with developing financial markets. They also report syster atic differences in the use of long-term debt between industrial and developing cou tries, as well as between small and large firms, even after controlling for the chara teristics of the enterprises.

The data in figure 1 show that firms in industrial countries clearly have mo long-term debt as a proportion of total assets. For example, the long-term debt-t asset ratio of an average firm in Norway, with a per capita gross domestic produ of $20,000, is five times greater than it is in Zimbabwe, with a per capita GDP less than $1,000. And large firms have more long-term debt as a proportion total assets and debt than do smaller firms (figure 2).

Figure 1. *Average Long Term Debt as a Percentage of Total Assets, 1980–91*

Source: Demirgüç-Kunt and Maksimovic (1996b). The data set consists of financial statement data for largest publicly traded corporations in manufacturing.

Importantly, this lack of long-term finance in developing countries persists even after controlling for firm characteristics and macroeconomic factors. Demirgüç-Kunt and Maksimovic explain this differential by institutional differences, such as the extent of government subsidies, different levels of development of stock markets and banks, and differences in the underlying legal infrastructure. Their results indicate that although policies that help develop legal and financial infrastructure are effective in increasing access to long-term debt, different policies are required to lengthen the debt maturity of large and small firms. Improvements in the legal system, for example, benefit all firms, although the effect is less significant for the smallest firms, which have limited access to the legal system. In the study, legal effectiveness is measured on a six-point scale every year, so, for example, the United States has a value of 6 and India a value of less than 2 for the whole sample period. The results indicate that a 10 percent improvement in the index of legal effectiveness would increase the largest firms' ratios of long-term debt to total debt by more than 5

Figure 2. *Debt Ratios of Small vs. Large Firms for 30 Countries*

Note: The figure presents the average debt ratio across 30 countries by firm size. The firms in each country are divided into quartiles by value of total assets, and the average debt ratios of each quartile, calculated across countries, is reported. Countries in the sample are Australia, Austria, Belgium, Brazil, Canada, Germany, Finland, France, Hong Kong, India, Italy, Jordan, Japan, Republic of Korea, Malaysia, Mexico, Netherlands, Norway, New Zealand, Pakistan, Singapore, South Africa, Spain, Switzerland, Sweden, Thailand, Turkey, United Kingdom, United States and Zimbabwe.

Source: Demirgüç-Kunt and Maksimovic (1996b).

percent. Thus, keeping all else constant, if a large Indian firm were to be transplanted to the United States, its ratio of long-term debt to total debt would increase by more than 100 percent. For the smallest firms, a 10 percent improvement in the legal index brings about an increase of 4 percent in their long-term debt ratio. In addition to being smaller in magnitude, this effect is also less significant in statistical terms.

Moreover, the authors find that policies to improve the functioning and liquidity of stock markets would benefit primarily large firms. During the sample period, the United States had one of the most liquid stock markets, with a turnover ratio of 60 percent, while Pakistan had one of the least liquid, with a turnover ratio of 11 percent. (Turnover is the value of shares traded divided by market capitalization.) The results indicate that a 10 percent increase in stock market liquidity leads to less than a 1 percent increase in the ratio of long-term debt to total debt for the largest firms. Holding everything else constant, however, transplanting the largest Pakistani firm to the United States would increase the ratio of the firm's long-term debt to total debt by 40 percent. This effect is not significant for the smallest firms.

In contrast, policies to improve the banking system would give smaller firms better access to long-term credit. The authors use the ratio of bank assets to gross domestic product as an indicator of banking development. In the sample, Germany has one of the largest banking systems—with a ratio of more than 100 percent—and Zimbabwe one of the smallest, with a ratio of 17 percent. The results indicate that a 10 percent increase in banking size leads to about a 3 percent increase in the ratios of long-term debt to total debt of the smallest firms. And again, keeping everything else constant, a small Zimbabwean firm transplanted to Germany would increase its long-term debt ratio by more than 100 percent. This result is not significant for large firms.

The Effect of Debt Maturity on Performance

In recent years theorists have been studying the forces that determine the maturity structure of a firm's debt (Brick and Ravid 1985; Diamond 1993; Kale and No 1990). This literature provides an interesting perspective on how this choice affects the enterprise's performance by emphasizing the different control and incentive properties of long- and short-term debt. In most of these models, long-term debt is not a necessity, but rather one of a number of financial claims that a firm may issue. Indeed, these models highlight several of the undesirable effects of long-term debt.

First, a capital structure that excludes long-term debt may be more efficient. Relying on such debt leads to greater distortions in the risk preferences of owners and managers, providing them with incentives to invest in projects that benefit themselves at the expense of outside investors (Myers 1977). This conflict can be mit

gated, however, by reducing the overall degree of leverage, or the maturity of debt, since the short maturity limits the period during which an opportunistic firm can exploit its creditors without being in default (Diamond 1991, 1993).

Second, short-term debt may also increase efficiency because of its role in disciplining management (Jensen 1986). Because of the more continuous scrutiny of a firm's operations and the threat of liquidation, short-term debt may in fact constrain wasteful activities.

Third, long-term debt allows management to delay its response to deteriorating market conditions and avoid exiting the market when it is obvious that the firm is failing. This reduction in efficiency is even worse if the loan is subsidized. In addition, when the market for refinancing short-term loans is competitive, long-term capital always increases the firm's financing costs and reduces the incentive to exert the effort to save on interest costs and increase efficiency (Rajan 1992; Ofek 1993).

Fourth, debt maturity is also correlated with credit quality and the profitability of existing projects. In the presence of asymmetric information about borrowers, firms of higher quality should choose short-term debt because they will be able to take advantage of the revelation of future good news (Diamond 1991). This positive information effect outweighs the liquidity risk of not being able to refinance oneself and running the risk of being liquidated by the lender. The opposite is true for firms with lower credit rating. Firms with the lowest credit ratings have access only to short-term debt, leading to a nonlinear relationship between maturity and credit rating.

But shorter debt maturity is not all good. Fear of liquidation may induce firms to avoid investing in profitable projects if the returns accrue in the distant future. Similarly, they may be reluctant to adopt more productive technologies unless they provide an immediate payoff. This shortening of the investment horizon may have negative consequences on overall performance. The faster the return on investment, the shorter will be the optimal payment structure. This provides a rationale for firms with long-term assets to have a longer debt maturity structure (Hart and Moore 1995). If financial markets undersupply long-term credit because banks are unable to internalize the full benefits of monitoring the firm (Mayer 1989; Calomiris and Himmelberg 1993) or because few people participate in financial markets (Diamond 1996), firms with a longer asset maturity may be disadvantaged.

On balance, the theoretical literature is inconclusive on how the maturity of debt affects a firm's performance. Notwithstanding data problems, empirical analysis in this area is difficult because it is not appropriate to draw conclusions about performance by simply treating maturity structure variables as independent, given that theory says that expected growth and profitability also affect the choice of maturity. The recent empirical literature attempts to avoid this simultaneity problem by focusing on performance indicators, such as efficiency measures, that should not play a role in "causing" maturity choice, or by using in-

struments for maturity choice, such as legal efficiency indicators, that measure the ability to enter into long-term contracts. This literature provides some interesting answers.

Evidence from Country Cases

Most of the empirical work in this area has been on growth and external financing. In a seminal work on the effect of financial constraints on investment decisions, Fazzari, Hubbard, and Petersen (1988) show that investment by U.S. firms is sensitive to cash flow. In later works, Calomiris and Hubbard (1995), Calomiris, Himmelberg, and Wachtel (1995), Carpenter, Fazzari, and Petersen (1994), and Calomiris and Himmelberg (1996) argue that the high shadow cost of external finance will show itself most clearly in the cash flow sensitivity of inventories. But surprisingly little work has been done on the links between debt maturity and performance. Gilson, John, and Lang (1990) find that the more long-term debt a firm has, the more likely it will be to reorganize successfully. Hall (1992) reports that when the ratio of long-term debt to physical capital increases, physical investment and research and development expenditures contract. Atiyas (1991), who investigates directed credit programs in Colombia, reports a negative relationship between long-term indebtedness and efficiency.

More recently, a number of case studies were conducted in industrial and developing countries using information about firms in each country: Schiantarelli and Sembenelli (1996) on the United Kingdom and Italy; Jaramillo and Schiantarelli (1996) on Ecuador; Calomiris, Halouva, and Ospina (1996) on Colombia; and Schiantarelli and Srivastava (1996) on India. In general, these studies find that short-term debt has no effect on efficiency and growth. The conventional wisdom that more long-term debt may actually lead to productivity improvements was confirmed in Ecuador, Italy, and the United Kingdom. Echoing the earlier findings for Colombia, however, the positive effect of long-term debt in Italy is substantially reduced—and even reversed—if the debt is subsidized (Schiantarelli and Sembenelli 1996). There is no evidence that the total investment is sensitive to the amount of long-term credit.

Evidence from Cross-Country Studies

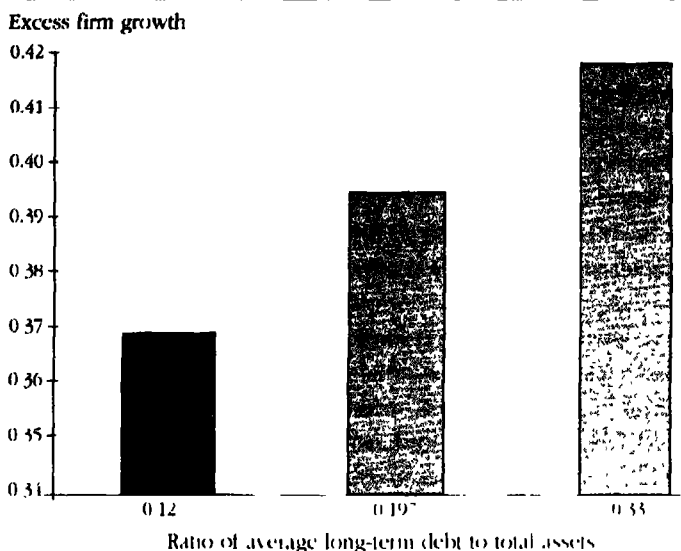
Rajan and Zingales (1996), who look at industry-level data across 27 countries, find a positive correlation between dependence on external financing and growth. They do not distinguish between external debt and equity, however, or between differences in debt maturity. Demirgüç-Kunt and Maksimovic (1996c) analyze firm-level data for 30 countries and find that an active stock market and the

ability to enter into long-term contracts enable firms to grow at faster rates than they could attain by relying on internal capital and short-term credit alone.

This research examines whether differences in financial and institutional development prevent firms in some countries from investing in potentially profitable growth opportunities. Such an effect, if it exists, would not affect all firms equally. Firms that can finance operations from retained earnings will be minimally affected, whereas firms whose financing needs exceed their internal resources may be severely constrained. To gauge the effect of financial development on the firm's ability to exploit growth options, it is necessary to identify firms that have a need for external financing and examine whether their realized growth depends on the development of financial markets.

To get around this problem, Demirgüç-Kunt and Maksimovic calculate growth rates of firms that rely only on internal financing. Their estimate of the constrained growth rate is based on the standard "percentage of sales" approach to financial planning (see Ross, Westerfield, and Jordan 1995). For each firm they define a maximum short term financed growth rate, which is an estimate of the maximum rate of growth if a firm reinvests all earnings and obtains enough short-term credit to maintain the ratio of its short-term borrowing to assets. They then calculate the proportion of firms that exceed their benchmark growth rates, average these over the 1980–90 period, and relate these excess growth rates to firm characteristics, including asset composition, profitability, and size; macroeconomic indicators, such as inflation and the growth rate of the economy; institutional factors, such as the effectiveness of the legal system and the level of government subsidies to enterprises; and financial indicators, such as banking development and stock market liquidity.

Their results show that the proportion of firms that grow faster than the predicted rate is related to specific features of the legal and financial systems and to the institutional structure of the economy. Even after controlling for firm characteristics, macroeconomic environment, financial development, and the extent of government intervention, the results show that firms with higher long-term debt ratios tend to grow faster than they would if they relied solely on internal resources (figure 3). Because it is difficult to identify the direction of causality in this framework, the authors also test their results by replacing the debt ratio with an index of legal effectiveness that they have shown to be highly correlated with long-term debt (Demirgüç-Kunt and Maksimovic 1996b). The results indicate that in countries whose legal systems score high on an efficiency index —which is expected to be an exogenous indicator of the ability to enter into long-term contracts—a greater proportion of firms grow at faster-than-predicted rates. They also find that an active, although not necessarily large, stock market is associated with faster growth. An additional test of causality using initial values of stock market liquidity and legal effectiveness to predict future growth does not alter the results. The findings

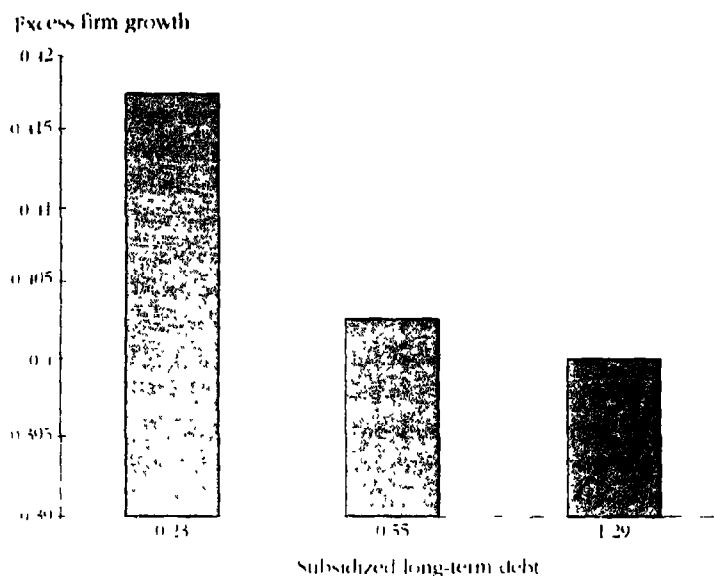
Figure 3. *Firm Growth and Long-Term Debt*

Note: The y-axis is the proportion of firms growing faster than predicted rates. Predicted rate is the rate which a firm can grow by only relying on retained earnings and short-term credit. Thirty countries in the sample were divided into three equal groups based on the average long-term debt to total assets [illegible] of firms. Countries with the lowest [illegible] are Brazil, Hong Kong, Jordan, Malaysia, Mexico, Pakistan, Thailand, Turkey, the United Kingdom, and Zimbabwe. Australia, Belgium, India, Italy, Japan, Korea, Netherlands, Singapore, South Africa, and Spain are in the second group. The third group comprises Austria, Canada, Finland, France, Germany, New Zealand, Norway, Switzerland, Sweden, and the United States.

Source: Calculations based on Demirgüç-Kunt and Maksimovic (1996c).

make it clear that the underdevelopment of financial markets and institutions prevent firms in developing countries from investing in potentially profitable growth opportunities.

Do policies intended to increase the availability of long-term debt actually work? Although government subsidies have increased the long-term indebtedness of firms around the world, Demirgüç-Kunt and Maksimovic (1996c) find no evidence that government subsidies to firms in the 30 countries in their sample were associated with faster growth (figure 4).

On the contrary, their evidence indicates that although long-term debt is associated with greater numbers of firms growing at rates that are higher than predicted, this result is reversed to the extent the credit is government-subsidized. Country case studies also confirm this effect. And Schiantarelli and Sembenelli (1996) show that the positive effect of debt maturity on firm performance declines as the aggregate proportion of subsidized credit rises—until it is reversed.

Figure 4. *The Effect of Subsidized Long-Term Debt on Enterprise Growth*

Note: The y axis is the proportion of firms growing faster than predicted rates. Predicted rate is the rate at which a firm can grow by only relying on retained earnings and short-term credit. Twenty-six countries for which data were available were divided into three groups (9, 9, 8) based on a ranking of the extent of subsidized long-term debt obtained by multiplying average LTD/TA ratios of firms by the ratio of government subsidy to gross domestic product. Countries with highest subsidies are Austria, Belgium, Brazil, Finland, India, Italy, Korea, Norway, and Sweden. The middle group includes Canada, France, Germany, Malaysia, Netherlands, New Zealand, Pakistan, Spain, and Switzerland. Finally, Australia, Japan, Mexico, Singapore, Thailand, Turkey, the United Kingdom, and the United States have the lowest amount of subsidized long-term debt.

Source: Calculations based on Demirgüç-Kunt and Maksimovic (1996c).

Lessons for Development Economists

First, even after adjusting for the characteristics of individual firms, long-term credit is scarce in developing countries, particularly for smaller firms. Thus if a firm in a developing country were magically relocated to an industrial economy, other things being equal, it could expect to receive more long-term credit. Second, among manufacturing firms, there is clear evidence that more long-term finance tends to be associated with high productivity. Third, and perhaps most important, this last result is reversed to the extent that such credit is subsidized.

Therefore, although it is worthwhile for governments to attempt to foster the supply of long-term credit, it is crucial that these interventions be crafted with great care—and little subsidy. If the macroeconomic environment is unstable, it

is unlikely that the market will provide long-term finance. When fixed interest rates are offered, savers regularly show that they are averse to putting their assets into long-term instruments; the yield curve is—or may soon become—inverted. When variable interest rates are offered, borrowers, or at least those who intend to repay, will not readily enter into a contract that could leave them bankrupt. Additionally, real sector reforms are important, as they will lead to changes in relative prices and in the performance of firms. If relative prices are clearly out of line, investors will not want to enter into long-term arrangements, so there will not be much demand for long-term credit. Beyond these reforms, however, firm characteristics do not admit to easy change. Institutional differences, such as the adequacy of banking and stock markets and the legal infrastructure, are important in affecting the supply of long-term credit and can be changed, although not overnight. Focusing attention on basic financial infrastructure is a "low distortion" road to achieving more long-term credit. Similarly, improving the legal system and collateral registration would especially redound to the benefit of smaller borrowers, for whom contract enforcement issues (and collateralized finance) tend to be of overriding importance.

Overwhelmingly, long-term finance tends to go to larger companies. That may be good for growth if there is evidence that larger firms tend to be more productive, but societies may be willing to sacrifice some growth for more equity. Unfortunately many schemes designed to facilitate access to credit for small firms have not achieved their goal. Research shows that banking and stock market development are complementary, probably because each produces and demands better information (Demirgüç-Kunt and Levine 1996). Hence, banking development will not only improve small firms' access but indirectly help larger firms by leading to greater capital market development as well.

The development of pension funds, insurance, and contractual savings systems are similarly low-distortion paths to stimulating long-term finance (Vittas 1996). Some economists argue that moving to a fully funded old age pension system would induce people to save more, but this has been a subject of considerable controversy on both theoretical and empirical grounds. In the case of Chile, however, a switch from a pay-as-you-go public pension plan to a fully funded private plan increased private savings substantially, although this is likely to be a transitory effect. Even in the absence of a permanent increase in savings, most analysts agree that the development of private contractual savings institutions leads to an increase in the supply of long-term funds because in every country these institutions hold a portfolio dominated by long-term assets. Thus, encouraging mandatory, fully funded pension schemes is an appealing way to encourage such credit indirectly. How rapidly governments will want to move in this direction depends on several variables, not least of which are demographic trends and the likelihood that investment pools can be allocated free of government interference.

Directions for Research

What do we still need to know about long-term credit? First, it is important to confirm the findings reported here. Although deep data sets that include listed and unlisted firms are still relatively rare, testing the findings here with new data as they become available would be appealing and worthwhile. Additionally, it is important to take a closer look at how countries have allocated such credit because it *can*, but does not always, have a positive effect on performance. Although government interventions have been costly and inefficient in many cases, credit policies in some countries have succeeded by establishing credible mechanisms that ensure proper allocation and repayment of funds (World Bank 1993). In-depth case studies of individual firms would help generate more precise recommendations about how credit policies can achieve optimal growth and equity outcomes; in several countries, long-term credit institutions face insolvency partly as a result of these programs. A careful balancing of the benefits and costs of credit-market interventions is therefore a top research priority.

Notes:

Gerard Caprio, Jr., is a research manager and Asli Demirgüç-Kunt is senior economist in the Development Economics Research Group of the World Bank. They would like to thank Stijn Claessens, Harry Huizinga, Vojislav Maksimovic, Fabio Schiantarelli, and Mary Shirley for helpful comments.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Atiyas, Izak. 1991. "Efficiency, Corporate Indebtedness, and Directed Credit in Colombia." Working Paper 54. World Bank, Industry and Energy Department, Washington, D.C. Processed.

Barclay, Michael J., and Clifford W. Smith, Jr. 1995. "The Maturity of Corporate Debt." *Journal of Finance* 50:609–31.

Berglöf, Erik, and Ernst-Ludwig von Thadden. 1993. "Capital Structure with Multiple Investors." *Quarterly Journal of Economics* 109:1055–84.

Bolton, Patrick, and David S. Scharfstein. 1993. "Optimal Debt Structure with Multiple Creditors." LSE Financial Markets Group Discussion Paper 161. London School of Economics and Political Science, London. Processed.

Brick, Ivan, and Abraham Ravid. 1985. "On the Relevance of Debt Maturity Structure." *Journal of Finance* 40 (5):1423–37.

Brock, Philip. 1995. "High Interest Rates and Bank Recapitalizations." Policy Research Working Paper 1683. World Bank, Policy Research Department, Washington, D.C. Processed.

Calomiris, Charles. 1993. "Corporate Finance Benefits from Universal Banking: Germany and the United States, 1870–1914." NBER Working Paper 4408. National Bureau of Economic Research, Cambridge, Mass. Processed.

Calomiris, Charles, Thekla Halouva, and Sandra Ospina. 1996. "Debt Maturity, Financial As Accumulation, and the Cost of External Finance: Evidence from the United States and Colo bia." World Bank, Policy Research Department, Washington, D.C. Processed.

Calomiris, Charles, and Charles P. Himmelberg. 1993. "Directed Credit Programs for Agricult and Industry: Arguments from Theory and Fact." *Proceedings of the World Bank Annual Conf ence on Development Economics 1993.* Washington, D.C.: World Bank.

———. 1996. "Investment Banking Costs as a Measure of the Costs of External Finance." Wo ing paper. Columbia University, Graduate School of Business, New York. Processed.

Calomiris, Charles, Charles P. Himmelberg, and Paul Wachtel. 1995. "Commercial Paper, Cor rate Finance, and the Business Cycle: A Microeconomic Perspective." Working paper EC- New York University, Stern School of Business, New York. Processed.

Calomiris, Charles, and Glenn R. Hubbard. 1995. "Internal Finance and Investment: Evide from the Undistributed Profits Tax of 1936–1937." *Journal of Business* 68:443–82.

Carpenter, Robert, Steven M. Fazzari, and Bruce C. Petersen. 1994. "Inventory (Dis)Investme Internal Finance Fluctuations, and the Business Cycle." *Economic Activity* 2:75–138.

Demirgüç-Kunt, Asli, and Vojislav Maksimovic. 1994. "Capital Structures in Developing Cou tries: Evidence from Ten Country Cases." Policy Research Working Paper 1320. World Ba Policy Research Department, Washington, D.C. Processed.

———. 1996a. "Stock Market Development and Firm Financing Choices." *The World Bank E nomic Review* 10(2):341–69.

———. 1996b. "Institutions, Financial Markets, and Firm Debt Maturity." Policy Research Work Paper 1686. World Bank, Policy Research Department, Washington, D.C. Processed.

———. 1996c. "Financial Constraints, Uses of Funds, and Firm Growth: An International Co parison." Policy Research Working Paper 1671. World Bank, Policy Research Departme Washington, D.C. Processed.

Demirgüç-Kunt, Asli, and Ross Levine. 1996. "Stock Market Development and Financial Interm diaries: Stylized Facts." *The World Bank Economic Review* 10(2):291–321.

Diamond, Douglas W. 1984. "Financial Intermediation and Delegated Monitoring." *Review Economic Studies* 51:393–414.

———. 1991. "Debt Maturity and Liquidity Risk." *Quarterly Journal of Economics* 106:709–3

———. 1993. "Seniority and Maturity of Debt Contracts." *Journal of Financial Economics* 33:341–68.

———. 1996. "Liquidity, Banks, and Markets." Policy Research Working Paper 1566. World Ba Policy Research Department, Washington, D.C. Processed.

Fazzari, Steven M., Robert G. Hubbard, and Bruce C. Petersen. 1988. "Financing Constraints in Corporate Investment." *Brookings Papers on Economic Activity* 1:141–206.

Gilson, Stuart, Kose John, and Larry Lang. 1990. "Troubled Debt Restructurings: An Empiri Study of Private Reorganization of Firms in Default." *Journal of Financial Economics* 27:315–

Hall, Bronwyn. 1992. "Investment and Research and Development at the Firm Level: Does th Source of Financing Matter?" NBER Working Paper 4096. National Bureau of Economic Re search, Cambridge, Mass. Processed.

Harris, Milton, and Artur Raviv. 1990. "Capital Structure and the Informational Role of Debt *Journal of Finance* 45:321–49.

Hart, Oliver, and John Moore. 1995. "Debt and Seniority: An Analysis of Hard Claims in Co straining Management." *American Economic Review* 85:567–87.

Hoshi, Takeo, Anil Kashyap, and David Scharfstein. 1990. "The Role of Banks in Reducing th Costs of Financial Distress in Japan." *Journal of Financial Economics* 27:67–88.

Jaramillo, Fidel, and Fabio Schiantarelli. 1996. "Long-Term Debt in Ecuador: Access and Effect on Firms' Performance: Lessons from Ecuador." Policy Research Working Paper 1725. World Bank, Policy Research Department, Washington, D.C. Processed.

Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review* 76:323–39.

Kale, Jayant, and Thomas H. Noe. 1990. "Risky Debt Maturity Choice in a Sequential Game Equilibrium." *Journal of Financial Research* 13:155–65.

Mayer, Colin P. 1989. "Myths of the West: Lessons from Developed Countries for Development Finance." Policy Research Working Paper 301. World Bank, Policy Research Department, Washington, D.C. Processed.

Modigliani, Franco, and Marcus Miller. 1958. "The Cost of Capital, Corporation Finance, and the Theory of Investment." *American Economic Review* 48:261–97.

Myers, Stewart C. 1977. "Determinants of Corporate Borrowing." *Journal of Financial Economics* 5:147–75.

Ofek, Eli. 1993. "Capital Structure and Firm Response to Poor Performance: An Empirical Analysis." *Journal of Financial Economics* 34:3–30.

Rajan, Raghuram G. 1992. "Insiders and Outsiders: The Choice between Informed and Arm's Length Debt." *Journal of Finance* 47:1367–1400.

Rajan, Raghuram G., and Luigi Zingales. 1995. "Is There an Optimal Capital Structure? Evidence from International Data." *Journal of Finance* 50:1421–60.

———. 1996. "Financial Dependence and Growth." NBER Working Paper 5758. National Bureau of Economic Research, Cambridge, Mass. Processed.

Ross, Stephen A., Randolph W. Westerfield, and Bradford D. Jordan. 1995. *Fundamentals of Corporate Finance*. Chicago: Richard D. Irwin.

Schiantarelli, Fabio, and Alessandro Sembenelli. 1996. "Form of Ownership and Financial Constraints." Policy Research Working Paper 1629. World Bank, Policy Research Department, Washington, D.C. Processed.

Schiantarelli, Fabio, and Vivek Srivastava. 1996. "Debt Maturity and Firm Performance: A Panel Study of Indian Public Limited Companies." Policy Research Working Paper 1724. World Bank, Policy Research Department, Washington, D.C. Processed.

Siraj, Khalid. 1983. "Report of the Task Force on Portfolio Problems on Development Finance Companies." World Bank, Washington, D.C. Processed.

Stiglitz, Joseph E., and Marilou Uy. 1996. "Financial Markets, Public Policy, and the East Asian Miracle." *The World Bank Research Observer* 11(2):249–76.

Stohs, Mark H., and David C. Mauer. 1996. "The Determinants of Corporate Debt Maturity Structure." *Journal of Business* 69:279–311.

Titman, Sheridan, and Roberto Wessels. 1988. "The Determinants of Capital Structure Choice." *Journal of Finance* 43:1–19.

Vittas, Dimitri. 1996. "Contractual Savings and Term Finance." World Bank, Policy Research Department, Washington, D.C. Processed.

World Bank. 1989. "Report of the Task Force on Financial Sector Operations." Financial Sector Development Department, Washington, D.C. Processed.

———. 1993. *The East Asian Miracle: Economic Growth and Public Policy*. New York: Oxford University Press.

From Prices to Incomes: Agricultural Subsidization without Protection?

John Baffes • Jacob Meerman

Drawing from the experience of the direct income support programs recently introduced in the European Union, Mexico, and the United States, this article highlights problems that may arise when the agricultural sector of a developing economy moves from price-based subsidization to less distorted income support. Such programs are a step in the right direction, but as currently implemented, they have many shortcomings. Moreover, developing countries may lack the necessary supporting arrangements needed to make such programs effective. The article argues that the programs should not restrict the use of land, that the programs should last for a stipulated period of time, and that the fiscal costs should be contained by linking income support payments to world prices.

Interventions in pricing policies play a vital role in the performance of agriculture, but they have often had an unfavorable effect on economic development (World Bank 1986; Lvers and Anderson 1992; Krueger, Schiff, and Valdés 1992; Meerman 1997). In industrial countries the allocative inefficiency resulting from the pricing policies of agricultural protectionism puts the deadweight losses to a protectionist government somewhere between 0.5 percent and 3 percent of gross domestic product (GDP) (Burniaux and others 1990). Fiscal costs have also been high and are increasingly viewed as excessive by predominantly urban electorates.

In response, many countries have undertaken structural reforms to stimulate production by liberalizing farm prices and integrating them with those of the world economy. Moreover, responding to commitments under the Uruguay Round of the General Agreement on Tariffs and Trade (GATT) as well as other multilateral and regional arrangements, more countries are expected to embark on similar price and trade reforms. For example, developing countries are committed to reducing average price supports by 13 percent in 10 years (Valdés and McCalla 1997).

The move to eliminate distortions, a step that inevitably implies some redistribution of income, is bound to be resisted by those groups that are penalized. Producers, for instance, will welcome the removal of a commodity tax, but the treasury and

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 191–211

other groups that have benefited from the tax are likely to oppose it. Farmers will object to proposals to reduce producer subsidies, because the additional costs that result will threaten the profitability of agricultural production. (Winters 1987, 1989–90, and Gardner 1987, 1990, discuss the political economy of agricultural protection.)

To check such adamant political opposition, some governments have attempted to replace price-distorting subsidies with direct income support mechanisms so that current production decisions are independent of—or less dependent on support prices for specific crops. Because income support programs compensate farmers for the loss of income, they are politically feasible and may make reform easier to accept.

Whether such schemes are in fact effective mechanisms for eliminating price supports is of interest to developing countries for at least two reasons. First, contrary to the widespread belief that industrial countries subsidize agriculture while developing ones tax it, today many developing countries subsidize agriculture—or its specific subsectors. Second, price supports in industrial countries affect developing countries in several ways. Subsidy-induced surpluses in member countries of the Organisation for Economic Co-operation and Development (OECD) depress world prices and reduce the income derived from competing exports from developing nations. The direct losses to foreign producers resulting from U.S. farm programs, for example, are estimated at about $1 billion in 1990. This figure is much higher if one considers the deadweight losses. OECD countries protect agriculture to insulate their markets from external shocks, but such subsidies depress world prices and increase their volatility, thus increasing the pressure for commodity stabilization programs in developing countries. Estimates suggest that the operation of the European Union's Common Agricultural Policy (CAP) has doubled the variability of world dairy prices while raising prices of wheat and beef by almost 50 percent within the Common Market (Knudsen and others 1990).

Agricultural Protection in Developing Countries

Agricultural protection in developing countries is not uncommon. In a study of eight Latin American countries, Valdés (1996) found that Chile, Colombia, and the Dominican Republic protect agriculture, while Brazil offers minimal protection. Argentina, Ecuador, Paraguay, and Uruguay protect certain subsectors. From 1989 to 1992 Nigeria maintained official prices on wheat and coarse grains at an average of 82 percent and 92 percent, respectively, above their world market prices. Algeria, Morocco, and Tunisia protected the same commodities at a combined average of 34 percent and 13 percent. Mexico's effective tariff equivalents for wheat and coarse grains were 55 percent and 69 percent of world prices in 1989–93 (Ingco 1995).

Extremely high production-supporting subsidies in the agricultural sectors of former centrally planned economies were (and in several cases still are) the rule (Brooks 1993). With the exception of some countries of the former Soviet Union, developing countries that have protected agriculture have usually been importers of the crops protected. In these countries protection has generally taken the form of import restrictions, including tariffs, import prohibitions, or restrictive import licensing. Because the transfer was made directly from consumers to producers, protection in these agricultural importing countries did not require large budget outlays. (The standard analysis of the efficiency, distributive, and fiscal effects of an import tariff is provided in appendix A.)

But in the OECD countries, agricultural protection has usually taken the form of crop prices pegged above world-market levels. This form of protection requires substantial fiscal outlays to finance the subsidies needed to encourage exports of the crop surpluses induced by the high prices. By 1991 the European Union was allocating nearly 1 percent of its GDP (58 billion ECUs) to agriculture, most of it to support the extreme price distortions under the CAP. For example, the wholesale prices of sugar, rice, and butter in 1991 were 308 percent, 171 percent, and 247 percent of respective world market prices (Atkin 1993).

Inputs have been subsidized as well. Governments have held down prices of fertilizer, irrigation, seeds, electricity, credit, and insurance. Knudsen and others (1990) report that in the early and mid-1980s, fertilizer subsidies in Sri Lanka and Turkey cost about 1 percent of total GDP, while those in Côte d'Ivoire, Egypt, Gambia, and Tanzania ranged from 50 percent to 100 percent of the market price of fertilizer.

Who Benefits from Agricultural Protection?

Replacing price supports with direct income support to farmers not only affects the income distribution of the concerned groups (the equity effect), but also generates efficiency gains (the efficiency effect). Whether income should be redistributed (through price supports, tariffs, taxes, lump-sum transfers, and so on) is an equity question in the sense that the marginal welfare of the proposed benefiting group (producers) is in some sense valued more than that of the proposed losing group (consumers or taxpayers). In terms of the demand-supply diagram shown in the appendix (figure A.1), the introduction of a protective tariff on imported agricultural commodities implies that certain producers are valued more than consumers. The efficiency argument, on the other hand, centers on the question of eliminating deadweight losses, that is, increasing total welfare by reallocating capital and labor away from the protected sector (agriculture) to more productive uses. In the real world the question of eliminating agricultural protection must simultaneously address both equity and efficiency.

In theory, such programs can be Pareto-optimal while maximizing GDP. A Pareto optimal policy change is one in which no one is worse off as a result of the change. In the present context achieving such an outcome essentially requires tradeoffs. Those benefiting from the reduction in prices of agricultural commodities due to policy changes give up enough of their increased income to farmers to fully compensate the latter for the loss of income resulting from the policy change. Therefore, replacing price supports with a lump-sum income transfer to farmers can be an attractive policy option. Ideally, producers receive about the same income as before; the treasury is no worse off because it does not spend any more than it did before; and the economy as a whole is better off because resources are allocated more productively. As Gardner (1990:190) puts it: "The existence of deadweight losses from commodity market intervention implies that losers should be able to compensate the gainers a bribe that exceeds their surplus gains, while the losers are better off paying the bribe than en during the intervention. The maximum size of the net gain is the deadweight-loss triangle. The reasoning, based on the compensation principles mentioned earlier suggests lump-sum transfers as a policy reform that provides a Pareto improvement The appendix provides a geometric version of this statement.

As a practical matter, whatever the aggregate net impact of the change from price to income support, these programs will also redistribute income within the farm community. Table 1 illustrates this point. The table is based on the Mexican proposal of 1994. In that year, Mexico's price supports on maize averaged about $59 a ton. Average annual national yields per hectare ranged from 1.60 tons (1989) to 2.35 tons (1992). The benefits to farmers were directly related to yields and the amount of land in maize (see the top panel of table 1). The second panel shows the benefits that would have resulted from the proposed plan to provide income support payments of $113 a hectare, up to a ceiling of 100 hectares. The "indifference" yield (that is, the point at which producers would be indifferent between price or income support) was 1.90 tons a hectare, which was also the median yield for 1989–92. The payment at the indifference yield was designed to compensate exactly the loss from eliminating price support. The big losers under the proposal would be high yield farmers cultivating more than 100 hectares. Subsistence farmers would be the (relative) big winners. To the degree that they consumed the maize they produced, they received no payments under the price support regime. Table 1 shows this effect by assuming that the first hectare of production is not sold but consumed.

Reforming Agricultural Support

The European Union, Mexico, and the United States have all taken recent steps to reform their agricultural price subsidies. How have these policy changes affected equity and efficiency, and what lessons do these reforms have for developing countries?

able 1. ***Hypothetical Transfers to Producers under Price and Income Support***
ollars)

Hectares	*1.60 tons/hectare*	*1.90 tons/hectare*	*2.35 tons/hectare*
	Assuming price support of $59/ton with a yield of		
1	0	0	0
50	4,758	5,650	6,988
100	9,516	11,300	13,976
115	10,943	12,995	16,073
200	19,032	22,600	27,953
	Assuming income support of $113/hectare with a yield of		
1	113	113	113
50	5,650	5,650	5,650
100	11,300	11,300	11,300
115	11,300	11,300	11,300
200	11,300	11,300	11,300
	Gain or loss as a result of an income support program		
1	113	113	113
50	892	0	-1,338
100	1,784	0	-2,676
115	357	1,695	-4,773
200	7,732	11,300	16,653
	Per hectare gain or loss as a result of an income support program		
1	113	113	113
50	18	0	-27
100	18	0	-27
115	3	15	-42
200	39	57	-83

Note: It is assumed that production from the first hectare is consumed by the households. The data resemble the program in Mexico for maize, 100 hectares is the limit of the payment, $59/ton reflects the 1989 price support of maize, $113 was the originally projected per hectare PROCAMPO (Programa Nacional de Modernización del Campo) payment (before the 1994 devaluation). 1.60, 1.90, and 2.35 tons/hectare were the average national maize yields for 1989, the median of 1989–92, and 1991, respectively. At a yield of 1.90 tons a hectare, producers are indifferent between price and direct income support. This can be seen in the lower panel where producers up to 100 hectares register neither gains nor losses.

Source: Authors' calculations, see text.

European Union: The 1992 CAP Reform

The costs to the European Union (EU) of the Common Agricultural Policy are of three kinds: Consumers pay far more for food than they need to; the deadweight losses are high—perhaps as much as 3 percent of the EU's gross domestic product; and budget outlays are high. In recent years the Common Agricultural Policy cost the EU nearly 1 percent of member GDP. These budget costs have become politically

Table 2. ***Characteristics of Agricultural Support Programs in the European Union, Mexico, and the United States***

Characteristic	*EU: CAP reform implemented 1993*	*Mexico: PROCAMPO implemented 1994*	*U.S.: FAIR implemented 1996*
Objective	To compensate producers for a reduction in support prices	To compensate producers for the elimination of guarantee prices on support crops	To compensate producers for the elimination of deficiency payments
Payment basis	Average acreage in support crops during 1989–91	Average acreage in support crops during 1991–93	Acreage for which deficiency payments were received in any of the past 5 years
Supported products	Wheat, maize, barley, rye, oats, rapeseed, sunflower, soybeans, dried pulses, beans, tobacco, beef, lamb	Wheat, maize, sorghum, barley, rice, cotton, beans, soybeans, safflower	Wheat, maize, sorghum barley, rice, cotton, oats
Time profile	Fixed in nominal terms; no expiration date	Total of 15 years, first 10 years fixed in real terms; declining in final 5 years	Program lapses after 7 years unless extended
Payment limits	None	$6,700 per farm	$40,000 per farm
Restrictions on the use of support-crop land	Land should be allocated to support crops; large producers must put into fallow a predetermined level of support-crop land	Land should be allocated to support crops but since 1996 land can be allocated to other agricultural uses	Land should be kept in agricultural uses (excluding fruits and vegetables), use must be in compliance with existing conservation plans
Other features	Support prices continue for cereals at lower level	"Negotiated" prices in effect for the first 2 transition years of the program; floor prices are retained for maize and beans	Nonrecourse government guaranteed commodity loans are retained in modified form

Note: The upper limit for PROCAMPO payments is 100 hectares and the per hectare payment is currently [illegible] Mexican pesos, or approximately US$6,700 (at 7.2 pesos/US$). Following the 1994 devaluation, PROCAMPO payments were not fully adjusted to inflation.

Source: Commission of the European Communities (1995); USDA (1996); SARH (no date).

intolerable and have been by far the strongest impetus for reform. In 1993 the EU reduced support prices on grains, oilseeds, and pulses and began to compensate producers by direct payments—based on past acreage in these crops—in conjunction with measures limiting the acreage for current production (table 2). Since that time, producer prices for these crops have declined by one-third.

Mexico: The 1994 PROCAMPO Program

Traditionally, CONASUPO (Compania Nacional de Subsistencias Populares), Mexico's agricultural parastatal, has been heavily involved in marketing, transporting, storing, and processing most of the country's agricultural commodities. Maize, beans, and wheat, by far the most important agricultural commodities, have been heavily subsidized through a system of guaranteed producer prices. The government also set panterritorial and pan-seasonal prices, which were usually announced before planting decisions were made. CONASUPO bought unlimited quantities at the guaranteed prices. Because producers knew in advance the price they were going to receive, they could shift production to those crops with the highest degree of relative protection rather than to crops that were the most profitable on the world market. The poorest peasants did not benefit from guaranteed prices because they hardly produced for the market.

In 1994 Mexico introduced a new farm program, PROCAMPO (Programa Nacional de Modernizacion del Campo), to provide income support to grain and oilseed producers—about 90 percent of all Mexican farmers (World Bank 1995; SARH no date). Under this regime prices of the nine crops in the program have become—in law at least—market driven or autonomous. Thus production and trade should become less distorted. Moreover, PROCAMPO is distributionally more attractive than the earlier guaranteed price support because poor subsistence farmers are eligible for payments and there is a ceiling of 100 hectares on the amount of land that any single farmer can claim to justify payments (see table 2, box 1).

The United States: The 1996 FAIR Act

Before 1996 the U.S. government compensated farmers who participated in wheat, feed grains, rice, and cotton programs through a system of so-called deficiency payments. The payment rate was based on the difference between the target price (set by the government) and the higher of the market price or the price at which the government would value crops used as collateral for loans made by the Commodity Credit Corporation, a public entity. The total payment was equal to the payment rate, multiplied by a farm's eligible payment acreage (usually the amount of land devoted to cultivation of the crop in question) and the program yield (which was established

Box 1. *A Closer Look at Mexico's Program*

Mexico was not adequately prepared to implement its new income support scheme. Several polic implications can be derived from its experience.

First, the program was announced well in advance of the registration date for eligible producer The lag allowed many farmers to increase the amount of land used for production of the eligib commodities and thus to increase their future payments. So rather than moving resources to mo efficient uses, the scheme was manipulated, initially at least, to move resources into production th was already inefficient. Moreover, because land rights among landowners, tenants, and sharecroppe were unclear, it has been difficult to determine who was entitled to the payment.

Government credibility was also an issue. Initially, some producers did not believe that th government would actually implement the program. Fearing increased taxation, they underreporte the amount of land allocated to eligible commodities. Furthermore, the fact that PROCAMPO initiall delinked payments from the current use of land but later required that the land continue to b allocated to the eligible crops may have further discredited the government. (In 1996 the governme increased the number of eligible crops.)

The macroeconomic environment also played an important role. Before the 1994 devaluatio most grains were highly protected through import restrictions. The devaluation sharply reduce protection relative to world prices. Although prices of tradable inputs also increased, it can be argu that to a large extent the devaluation substituted for PROCAMPO.

Market performance is also relevant. Mexico's old system of guaranteed prices effectively preclude any kind of autonomous market. Thus markets are still underdeveloped and inefficient. Movin from guaranteed prices to a system of income support is more complicated in Mexico than in countr with a strong tradition of competitive commodity markets.

for the particular farm by the Department of Agriculture on the basis of historic yields).

On April 4, 1996, the Federal Agricultural Improvement and Reform Act (FAIR) became law, after the longest debate on a farm bill in U.S. congressional history (USDA 1996). FAIR removed the link between income support payments and farm prices by providing "production flexibility contract payments" for several crops Participant producers receive these payments as a function of the amount of lan registered for government support payments in earlier years. The payments are independent of current production, and farmers therefore have a more flexibl incentive structure regarding planting decisions. The payments are fixed annuall at a declining rate and under current legislation will end after seven years (see table 2). The practice of leaving land idle, required as part of the previous support pro grams, has been eliminated and producers are now free to plant any crops on the former contract acres except fruits and vegetables. The result is that producers depend more heavily on the market and also bear greater risk from increased price variability.

Assessing the Reform Programs for Efficiency and Equity

On efficiency grounds the three reform programs are improvements. All three are similar in delinking support from current production decisions and in moving toward market-driven prices. All three promote allocative efficiency in that quantity produced and price received are—in varying degrees—independent of the amount of support. Thus, the three income support schemes are among the less distortionary mechanisms for maintaining agricultural incomes above their market outcomes.

All three programs, however, are less than ideal in that the use of land is not delinked from the program. This requirement probably reflects political considerations, as the payments must be seen to be going to "true" farmers. CAP reform requires either that land remains in production of the crops eligible for support or that producers set aside (hold out of production) a predetermined amount of support-crops land. PROCAMPO also holds land in agricultural production but permits a greater variety of crops to be cultivated. FAIR requires that land be kept in general agricultural use, but cannot be switched to fruits and vegetables, while producers must also comply with existing conservation plans. These restrictions obviously reduce allocative efficiency. In developing countries, where landownership records are poor, they may also pose implementation problems because monitoring compliance with cultivation restrictions would be difficult, especially where the average farm size is small. Monitoring such restrictions may also lead to corruption.

Do these reforms promote equity? Because the payments under the new programs depend on the amount of land used for past agricultural production, large farmers will receive the largest payments (although a limit on the total paid to any single farmer will reduce this effect). In industrial countries high or no payment ceilings have been more or less an acceptable practice, because most farmers have been perceived to have lower standards of living than other citizens. This perception is clearly the case in a number of OECD countries, where one of the most commonly declared objectives of agricultural policy has been a "satisfactory and equitable standard of living for farmers" (Winters 1989–90:241). This perception is probably decreasing. In the EU the belief is gaining ground that the "vast bulk of CAP money goes to farmers, many of whom are well off." (*Financial Times* 1997). In many developing countries (especially in South Asia and Latin America), a few people with relatively high incomes own most of the land; many farmers own little or no land. In these cases, where land is distributed very unequally, unlimited payments will normally be a less appropriate criterion.

But such programs may promote equity in those developing countries with a strong correlation between poverty and degree of subsistence production. Under such a program, poor subsistence farmers with land are better off because they can consume

the previously subsidized commodities and receive cash payments at the same time (see first row of table 1). Mexican farmers who own less than two hectares of land receive more than 8 percent of PROCAMPO payments, although they have historically marketed very little and therefore received little support under the old program because the price guarantees applied only to the traded portion of the commodity (Deininger and Heinegg 1995).

Fiscal Costs

The fiscal costs of income support programs can be problematic. Thus far, two of the programs have been more costly than the programs they replaced (at prices prevailing at the time of implementation). The cost of the U.S. program in 1996–97 was estimated at about $5.5 billion, as opposed to $4.2 billion in 1994–95 for deficiency payments (USDA 1996). In Mexico the introduction of PROCAMPO almost doubled the transfers to the agricultural sector—from 6.4 billion new pesos in 1993 to 11.7 billion new pesos in 1994 (SARH no date). Moreover, if the programs are expensive initially, their fiscal cost is expected to remain high because payouts are independent of world prices.

The costs of the new program compared with the old price support program depend on actual world prices. Under high world prices the transfers to farmers under price supports would be low—or would even cease. For example, in the United States, because of the 1996 boom in grain prices, support based on deficiency payments would have been $1 billion–$2 billion—a fraction of the $5.5 billion paid under the FAIR act. Under low world prices, price subsidies would be higher, making the new program a bargain.

Price and Other Risks

Replacing stable support prices or guaranteed prices with direct income transfers exposes producers to the risk of volatile market prices. The shock from the exposure to risk would be more prevalent in countries where the government assumed all price risks by maintaining panterritorial and panseasonal prices than in countries with other types of protection, such as subsidies per output unit, floor prices, import tariffs, and possibly nontariff barriers.

Short-term price volatility can be alleviated with devices that mitigate market-based risk, either through private initiative or with public assistance. Forward and futures markets are effective tools that can offer both price discovery and hedging not only to producers, but also to merchants or other concerned parties. Participation in existing futures markets by foreign producers and merchants, however, appears to be limited, especially in cases where the costs of production and marketing the commodity are predominantly in domestic currency (one

would have to hedge both the commodity as well as the exchange rate). Several developing countries are attempting to create their own futures exchange markets, however.

Government-assisted risk-mitigation devices are another option. In 1997 the U.S. Department of Agriculture introduced revenue insurance against both crop failure and falling prices. Initiated on a pilot basis in the states of Nebraska and Iowa, these revenue insurance policies have sold briskly (*Economist* 1997). Similarly, Mexico offers a guaranteed minimum price to cotton farmers for a predetermined fee through a government organization, ASERCA. The minimum price is based on the New York cotton futures exchange. ASERCA offers the guaranteed price in U.S. dollars and hedges the risk by using the predetermined fees to purchase put options on the exchange for future delivery after the harvest. Should prices fall, ASERCA pays the farmer the difference between the prevailing New York futures and the guarantee price. If prices rise instead, ASERCA makes no payment (Varangis and Larson 1996). Because the entire transaction is denominated in U.S. dollars, cotton producers assume the exchange risk. It should be emphasized, however, that private and government-assisted risk-mitigation devices deal with short-term price volatility, that is, one crop cycle at most.

Generally speaking, there are many ways to reduce risk in addition to formal measures. Farmers can grow a variety of crops with different market and climatic risks. Contract farming at fixed output prices eliminates price risk. An off-farm job for a member of the household is also a possibility.

Negative Supply Response

In theory, price and trade reforms can be Pareto improvements. They allow farmers to maintain their incomes at roughly the same levels while inducing growth in the economy through better resource allocation. Nevertheless, the removal of protectionist policies is likely to reduce agricultural output because lower crop prices reduce investment and accelerate the exit from agriculture. This has been the case in the former Soviet economies, frequently dramatically so. In Hungary, for example, market liberalization after 1989 reduced agricultural output by more than one third and induced the exit from agriculture of one-half the country's farmers.

On balance, the reforms undertaken by the Mexican government in the late 1980s and early 1990s (including the implementation of PROCAMPO) appear to have reduced the profitability of the crop sector (Baffes 1998; World Bank 1996). In the European Union, if recently suggested reforms to the CAP cut preferential assistance to oilseed producers, their income may drop by as much as 20 percent (*Oils & Fats International* 1997).

Moreover, contraction of the sector inevitably reduces the demand for farm labor, so the income of landless farm workers is likely to decline (at least in the short run), although they may benefit to the degree that the prices of food decline. The removal

of guaranteed prices in Mexico was initially expected to induce a long-term reduction of 6 to 24 percent in the agricultural labor force and a corresponding reduction in grain production of between 11 and 28 percent. (The estimates for the Mexican agricultural sector are from Baffes 1998; Levy and van Wijnbergen 1994; and Burfisher, Robinson, and Thierfelder 1992.) In response to the FAIR Act, U.S. farmers are already shifting land out of maize to soybeans, anticipating changes in the price structures of the two crops.

It is important to be aware of these consequences early on. Otherwise, when the reforms fail to induce a positive supply response or to increase the demand for labor at least in the short run, the confusion and resistance that is generated could jeopardize the reform process.

Toward a Successful Transition

Although all three programs move in the right direction, the considerations above suggest that they could have been designed and implemented more effectively. For one thing, the programs could have been more comprehensive. In addition reform programs should not restrict land uses, they should not cost more than the subsidization programs they replace, and they should be transitional.

Making Reforms Comprehensive

To realize the full benefits of an income support system, the programs should include all crops and substitute for all existing price support programs. FAIR does not apply to sugar, tobacco, peanuts, and milk, all of which are heavily protected in ways that seriously misallocate resources. And PROCAMPO is restricted to nine commodities; price floors are still in operation for the two basic crops (maize and beans), and several subsidy programs, especially on inputs, are still in place. Although CAP's reform covers most crops and some livestock, it fails to eliminate price supports and retains many quantitative restrictions.

Lifting Restrictions on Land Use

An income support program should not impose restrictions on the use of land. With the exception of environmental considerations, the main justification for restricting land use is to ensure that program payments go only to bona fide farmers. Yet one important reason for replacing subsidies with income support is to encourage individual producers to use resources as relative prices and comparative advantage dictate. Another is that future payments under the income support scheme could be

used as credit collateral or sold outright as an asset, facilitating investment in profitable activities. However, because it is unlikely that lending institutions would have the capacity (and authority) to verify program participation and program compliance, their value as collateral is limited.

Finally, enforcement of restrictions on land-use entails careful, and in some countries relatively expensive, administrative measures. In Mexico the majority of agricultural extension workers employed by PROCAMPO are engaged in monitoring the enforcement of restrictions on land use.[1]

Keeping the Costs Down

If they are to be politically feasible, income support payments must be designed so that they do not exceed the costs of the programs they replace. For this reason, governments should link payments inversely to world agricultural prices: when prices are high, producers receive lower income support payments and vice versa. Otherwise if world prices rise, producers not only receive these high prices but in addition are subsidized by the program. This imposes an additional burden on taxpayers who must pay high food prices and are also burdened with the program's costs.

In addition, rather than providing a uniform level of per hectare support, policymakers can set up a declining index with a ceiling. For example, the first 10 hectares receive full support, say $100 a hectare, the next 10 hectares receive less support ($50 a hectare), and so on, effectively increasing the relative support to small producers. If the program is announced before land registration is completed, however, farmers are likely to divide large holdings among family members, in effect neutralizing the effect of a declining index. Similarly, corporate farms can convert to holding companies for a large number of small farms. For the same reasons, such moral hazard also arises with programs that impose ceilings on benefits, such as PROCAMPO and FAIR (see table 2).

Nevertheless limits on the maximum acreage eligible for income support or simple benefit ceilings as in FAIR are equitable devices for containing fiscal costs. Such limits reduce support to larger producers and may promote more equal income distribution. Table 1 shows that such a ceiling can generate substantial savings by reducing the benefits to above-average-yield large producers.

Limiting the Duration of the Program

In view of the uncertainties, the programs should be transitional and not permanent. Although PROCAMPO is scheduled to be phased out in 15 years, at which time world prices will prevail, FAIR's language leaves open the question of support when the program expires after 7 years. Thus far, the CAP is not time-bound, although it is

likely to be subject to another round of reforms. Nevertheless, if the final object is—as it should be—to eliminate support to producers, the programs should wi down to an explicit expiration date.

It is important to distinguish between the short run and the steady state. Consic the Pareto-improvement element of such programs. Under the assumption that pr support programs would have continued indefinitely, an explicit time profile for t income support system implies that such programs are not necessarily Pareto i provements. This is the case because producers who previously were receiving su port may be worse off after the termination of the income program than they wou have been had price subsidization continued.

Establishing Supporting Institutions

The shortcomings discussed above apply to both industrial and developing cou tries. But weak supporting institutions are likely to be a problem in developi countries. To ensure that payments to producers are fair (that is, based on t amount of land farmed) and are paid in a timely manner, a national land regist must be developed before initiating the program—something that is extreme difficult to achieve in many countries, for example, those in tropical Africa. T government must also have policy credibility if producers are to react as desire Credibility was a problem in Mexico, where the amount of land in crops was fi under-reported in many areas (due to fear of government taxation), and then ove reported. As of 1993, the final determination of eligible areas had not yet o curred. In many instances, eligible areas were "negotiated" at the community lev (Salinger, Metzel, and Arndt 1995).

Clearly the macroeconomic environment, and particularly the exchange rate, shou be adequate and stable. In some cases eliminating currency overvaluation may ma it possible to eliminate protection without fiscal compensation. If substantial d valuation is likely, a feasible approach to eliminating price supports may be to ma explicit the benefits to farmers that result from devaluation and include eliminati of agricultural price support as part of the macroeconomic reform package. Finall efficient and integrated local commodity markets are needed to ensure a smoo transition.

Another set of problems stems from uncertain land tenure rights. *Who receives th payment* may be an issue, particularly in developing countries. Consider a case whe a tenant has cultivated the land for the entire period on which the payments a based. If support is based on land previously allocated to supported crops, the lan owner will claim the payments. On equally valid grounds, the tenant may claim th payments because it was his or her cultivation that qualifies the land for suppor Resolving this issue will often complicate the implementation process. (The Unite States has developed an elaborate legal definition of "producer"—including land

owner, share cropper, or rent-paying tenant—that stipulates the conditions of eligibility for receiving benefits under FAIR.)

In the absence of well-defined land tenure rights, appropriate preparation of a transition program may require an extensive survey of the rural economy to identify the distribution of farms operated by owners and those that are not. The survey will indicate the technical feasibility of the program but may also provide information on variables such as the distribution of farm size, yields, and commodities produced. Such information may be needed to design equitable and affordable programs that are politically acceptable; it will also indicate whether direct income support is an appropriate route to reform. These considerations suggest that in many—perhaps most—developing countries, informal land tenure arrangements may preclude subsidies based on land ownership mechanisms. As Binswanger and Deininger (1997:1966) have recently argued: "The costs of maintaining records, negotiating, contracting, and policing property rights can be high and may exceed the value of the land especially in rural areas with low population densities and little market access."

Concluding Remarks

In designing income support programs, it is well to keep in mind, first, that separating the payments from current production decisions implies consequences that are likely to be perceived as negative in many instances, even though economically rational outcomes are fully expected. Producers will frequently face higher risk from increased price variability. Because the ratio of output to input prices will be lower, agricultural output will decline for the crops in question, which in turn may reduce the demand for agricultural labor. As is the case with other types of support, the politically active large producers who will probably receive the lion's share of the support may not be those most in need of support during the transition period. The latter problem has been well recognized and is one of the complicating factors behind attempts to further reform the CAP.

Second, it is unlikely that the conditions and requirements discussed earlier will be fully met, particularly in the developing countries that protect agriculture. Decisionmakers need to be aware of these requirements to ensure that the programs are feasible, quite apart from the question of dealing successfully with the political forces that are involved.

Reforms, therefore, have to be considered within a broad policy perspective. Compensation could, for example, take the form of government-financed investments to benefit producers, such as increased expenditures for rural infrastructure. Most developing countries have an inadequate network of rural roads, inefficient water supply and irrigation networks, and limited electric power. Rural social ser-

vices are also neglected; education and health care in rural areas are inferior t those provided in cities. Supplying credit through group-based lending to micro enterprises in high population areas may be another appropriate instrument (Binswanger and Landell-Mills 1995). In Japan, for example, the Uruguay Round Agreement called for permitting imports of foreign rice equal to 4 percent to 8 percent of domestic consumption. Anticipating that this market opening would lead to lower domestic prices for rice and a decline in farmers' income, the Diet enacted law that appropriated more than 6 trillion yen (about $60 billion) for compensatory measures to increase farmers' incomes indirectly through improvement in infrastructure in rural areas, enhancement of agricultural technology, access to credit and so on (Goto 1997).

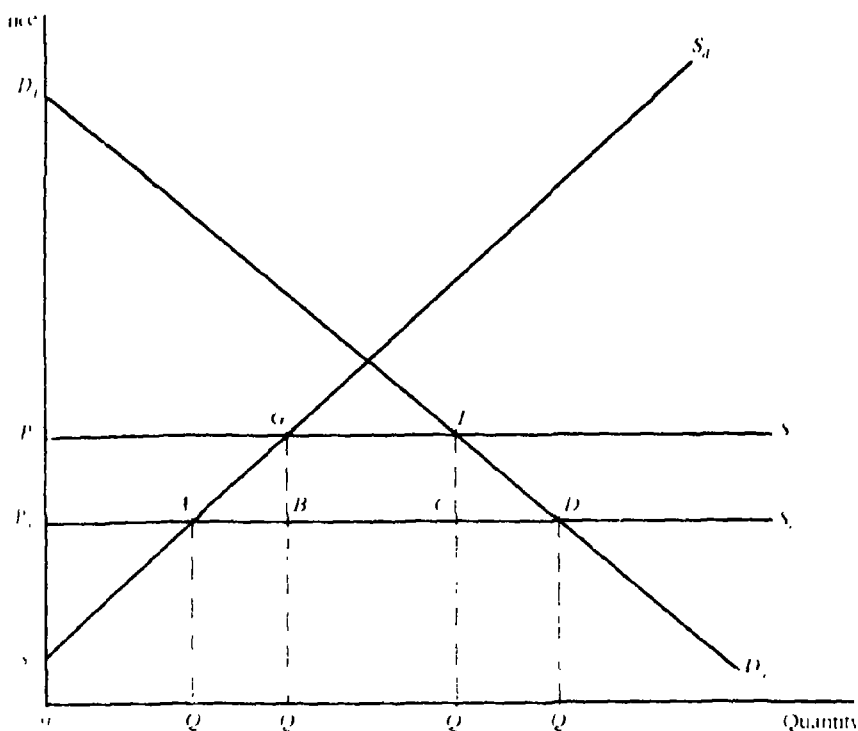
To conclude, it is important to recall that a direct income support program i intended to provide a transition from price-distorting subsidies to a liberalized sector that allows resources to be allocated more efficiently. It is not a poverty reduction program, although it can increase the income of subsistence landholders. It is not an investment program because it has no provisions on how the money is to be spent. I is not designed to induce agricultural growth because it lowers producer prices. Finally, because the program is linked to an asset—land—the lion's share of the payments may go to large producers.

Appendix A. Income Redistribution and Efficiency Losses under an Import Tariff

Let D_dD_d and S_dS_d denote domestic demand and supply of the commodity in question, while P_mS_m denotes the world price, which can also be viewed as the perfectly elastic supply of imports (figure A.1). In a closed economy the market clears at quantity and price dictated by the intersection of domestic demand and supply. If the economy is open, price is P_m, and the supply for the commodity becomes S_dAS. Domestic producers supply OQ_1 while the remaining (Q_1Q_0) is imported. This is the competitive outcome.

If the government introduces an import tariff t, the effective supply schedule becomes S_dGS_t. At P_t (the new price), consumers demand is Q_2. Domestic producers supply OQ_3, while the remaining (Q_3Q_2) is supplied by imports. An import tariff has several effects. First, consumers pay a higher price (from P_m to P_t) and consequently demand less (from OQ_0 to OQ_2). Second, domestic producers receive a higher price and hence produce more (from OQ_1 to OQ_3). Third, the government receives tariff revenue (the area $BGFC$). To summarize, consumers lose while producers and government gain.

The losses to the three groups, however, exceed the gains. The monetary burden on consumers (reduction in consumer surplus) due to the tariff is the area of P_mP_tH

igure A1. *Import Tariff in a Small Open Economy*

of which $BGFC$ (the tariff revenue) goes to the government; $P_w P_t GA$ (the change in producer surplus or change in profits) goes to the producers; the triangular area ABG (the excess cost of producing the additional output $Q_1 Q_3$) is the efficiency loss, that is the additional resources that the country uses to produce domestically the commodity that would have been saved, had the corresponding amount of the commodity been imported. Sometimes this area is termed deadweight loss, or Harberger triangle. Finally, the triangular area CFD denotes an additional welfare loss to consumers, or reduction in consumer surplus that is not offset by government revenue from the tariff. Total deadweight losses equal the sum of the two triangles.

After the introduction of the tariff, the relative size of producer gains and consumer losses depends on demand and supply elasticities and the level of the tariff. If the after-tariff price exceeds the price that prevails if the economy is closed, then imports are not taking place. In this case, the government receives no revenue, while the "transfers" take the form of increased producer profits and increased consumer welfare losses.

For agricultural producers the import-tariff setting is analogous to producer pric support with no production controls. A similar outcome arises if, instead of taxin imports, the country limits the imported quantity to, say, $Q_3 - Q_2$ The importe then receive a rent equivalent to the tariff revenue net of the cost of obtaining a import license. Under certain assumptions, importers competing for the import l censes would pay the same in the aggregate for licenses as the tariff forgone.

The fundamental idea behind a Pareto-optimal income support program that com pensates for tariff removal goes as follows. After eliminating the tariff, income i redistributed as lump-sum transfers. Producers plus importers (consumers) suppl (demand) OQ_0 at price P_m as in the no tariff scenario. Producers receive as compen sation for their losses the lump-sum transfer P_mP_tGA plus part of the two triangula areas and therefore are better off; the government receives $BGFC$ plus part of th triangular areas and therefore is better off; consumers pay P_mP_tFC and part of th two triangular areas but retain the remaining part of the two triangular areas an therefore are better off. Thus, from an efficiency point of view, a lump-sum transf can be a Pareto improvement.

Going from price to income support, therefore, can be a win-win move. Even i political forces aim to achieve some such outcome, the "transaction costs" involv in moving to the desired outcome may be high. First, consider the difficulties o measuring areas such as P_mP_tGA and ABG. Such measurement requires knowledg of the yield as a proxy for land-rent, that is, producer surplus, of each producer. It i difficult for governments to acquire such knowledge. Therefore the program mu be designed according to an "average yield," which, as table 1 indicates, will ove compensate some producers and undercompensate others. An additional difficul results from the choice of the base period. Because not all producers allocate th same amount of land to a particular crop each period, different periods will affe producers differently.

Second, there are the implications of the vertical linkages in the particular mark Reducing the price from P_t to P_m implies less output and therefore less use of inpu Because labor is in many instances a major input, the welfare of farm labor is lik to change, independently of what happens to landowners.

Third, there are horizontal linkages. The forces of complementarity and substitu ability among products will shift resources from one product market to another (de pending on relative prices and the structure of the respective technologies).

Finally, reforms frequently involve altering several policies simultaneously (su as price support, input subsidy, tariffs, and quantitative restrictions). The calculati of the associated benefits and losses by recipient group requires data that practica speaking are virtually impossible to collect.

In summary, the move from price to income support ideally must take into co sideration all the difficulties outlined above if the policy change is to be implemen with both increased efficiency and Pareto-optimality. Pareto-optimality as a goal

pful in structuring and evaluating proposed reforms. Its achievement is another ry.

)tes

n Baffes is with the Development Economics Research Group at the World Bank, and Jacob erman is an economist in the Operations Evaluation Department at the World Bank. The au- rs would like to thank Adolfo Brizzi, Louise Cord, Bruce Gardner, Junichi Goto, Tassos Haniotis, Alberto Valdés. The article has also greatly benefited from a discussion by participants at a rld Bank workshop on February 4, 1997.

Land tenure is one of the most difficult aspects of Mexican farm policy. In addition to private ns, there are quasi-communal farms, the *ejidos*. Created under the Constitution of 1917, *ejidos* ranteed all Mexicans land rights through expropriations of large landholdings. But *ejido* mem- s rights over land and water use were restricted. Sale or rental of *ejido* land was prohibited; mbers could not hire wage workers, and they could not be absent from their farm for more than years without losing their rights. The *ejido* system gave members little control in their choice of uts and outputs (Heath 1990). In 1992 a reform in the Mexican Constitution gave *ejidatarios* right to rent and sell land to outsiders with the approval of a majority of *ejido* members. Land now be pledged as a collateral. The titling process has been slow, however. By early 1995, only percent of *ejidos* had been given land titles (De Janvry and others 1995).

ferences

word "processed" describes informally reproduced works that may not be commonly available ugh library systems.

n, Michael. 1993. *Snouts in the Trough: European Farmers, the Common Agricultural Policy, and the Public Purse.* Cambridge, U.K.: Woodhood Publishers.

es, John. Forthcoming. "Structural Reforms and Price Liberalization in Mexican Agriculture." *Journal of International Development.*

swanger, Hans P., and Klaus Deininger. 1997. "Explaining Agricultural and Agrarian Policies in Developing Countries." *Journal of Economic Literature* 35:1958–2005.

swanger, Hans P., and Pierre Landell-Mills. 1995. *The World Bank's Strategy for Reducing Pov- erty and Hunger: A Report to the Development Community.* Environmentally Sustainable Develop- ment Studies and Monographs Series, No. 4. Washington, D.C.: World Bank.

oss, Karen M. 1993. "Challenges of Trade and Agricultural Development for East/Central Eu- rope and States of the Former USSR." *Agricultural Economics* 8:401–20.

rsher, Mary, Sherman Robinson, and Karen Thierfelder. 1992. "Agricultural and Food Policies in a United States–Mexico Free Trade Agreement." *North American Journal of Economics and Finance* 3:117–39.

rnaus, Jean Marc, John P. Martin, François Delorme, Ian Lienert, and Dominique van der Mensbrugghe. 1990. "Economy-Wide Effects of Agricultural Policies in OECD Countries: A Gen- eral Equilibrium Approach Using the Walras Model." In I. Goldin and O. Knudsen, eds., *Agricul- tural Trade Liberalization: Implications for Developing Countries.* Paris: OECD Development Center.

mmission of the European Community. 1993. *The New Regulation of the Agricultural Markets.* Luxembourg.

Deininger, Klaus, and Ayo Heinegg. 1995. "Rural Poverty in Mexico." Latin America and th Caribbean Regional Office, Country Department II, World Bank. Processsed.

De Janvry, Alain, Elisabeth Sadoulet, Benjamin Davis, and Gustavo Gordillo de Anda. 1995. "*Ejid* Sector Reforms: From Land Reform to Rural Development." Paper presented at the Conferen on Mexican Agrarian Reform, Columbia University, New York, April 6–7. Processed.

The Economist. 1997. "The Farmbelt Breaks Free." July 12, pp. 21–22.

Gardner, Bruce. 1987. "Causes of U.S. Farm Commodity Programs." *Journal of Political Econom* 95:290–310.

———. 1990. *The Economics of Agricultural Policies*. New York: McGraw Hill.

Goto, Junichi. 1997. "Regional Economic Integration and Agricultural Trade." Policy Researc Working Paper 1805. The World Bank, International Economics Department, Washingto D.C. Processed.

Heath, John. 1990. "Enhancing the Contribution of the Land Reform Sector to Mexican Agricu tural Development." Policy Research Working Paper 285 Latin America and the Caribbe Regional Office, Country Department II, World Bank, Washington, D.C Processed

Ingco, Merlinda D. 1995. "Agricultural Trade Liberalization in the Uruguay Round: One St Forward, One Step Back?" Policy Research Working Paper 1500. World Bank, Internatio Economics Department, Washington, D.C. Processed.

Knudsen Odin, John Nash, James Bovard, Bruce Gardner, and Alan I Winters. 1990. *Redefin the Role of Government in Agriculture in the 1990s* Discussion Paper 105. Washington D.C World Bank.

Krueger, A. O., M. Schiff, and A. Valdés 1992 *The Political Economy of Agricultural Price Interv tion in Latin America* San Francisco: International Center for Economic Growth

Levy, Santiago, and Sweder van Wijnbergen. 1994. "Labor Markets, Migration, and Welfare A culture in the North America Free Trade Agreement" *Journal of Development Econo* 43:263-78.

Meerman, Jacob. 1997. *Reforming Agriculture The World Bank Goes to the Market* Operations E ation Study. Washington, D.C.: World Bank

Oils and Fats International. 1997. "CAP Reform Plans Threaten Oilseed Production " August p

SARH (Secretaria de Agricultura and Recursos Hidraulicos) No date *PROCAMPO Vamos al Grano Progresar*. Mexico City.

Salinger, Lynn, Jeffrey Metzel, and Channing Arndt 1995 "Mexico. Devaluation and Maize M ket Reform." Report prepared for the Government of Mexico (Ministry of Agriculture L stock, and Rural Development), Mexico City, and the World Bank, Washington D.C Pro cessed.

Southey, Carolyn. 1997. "Seeds of Discount." *The Financial Times* Jan 24 15

Tyers, Rod, and Kym Anderson 1992. *Disarray in World Food Markets* New York Cambri University Press.

USDA (U.S. Department of Agriculture), Economic Research Service 1996 *Agricultural Ou Provisions of the 1996 Farm Bill*. Washington, D.C.

Valdés, Alberto. 1996. *Surveillance of Agricultural Price and Trade Policy in Latin America Major Policy Reforms*. World Bank Discussion Paper 349. Washington, D.C.

Valdés, Alberto, and Alex F. McCalla. 1997. "The Uruguay Round and Agricultural Polic Developing Countries and Economies in Transition." *Food Policy* 21:419-39.

Varangis, Panos, and Donald Larson. 1996. "Dealing with Commodity Price Uncertainty" Po Research Working Paper 1667. World Bank, International Economics Department, Wash ton, D.C. Processed.

Winters, Alan L. 1987. "The Political Economy of the Agricultural Policy of Industrial Countries." *European Review of Agricultural Economics* 14:285–304.

———. 1989–90. "The So-Called Non-Economic Objectives of Agricultural Support." *OECD Economic Studies* 13 (Winter):238–64.

World Bank. 1986. *World Development Report 1986.* New York: Oxford University Press.

———. 1995. "Mexico: Project Completion Report: Second Agricultural Sector Adjustment Loan." Report 14552. Latin America and the Caribbean Country Department II, Washinton, D.C. Processed.

———. 1996. "Mexico: Performance Audit Report: Mexico Agricultural Sector Adjustment Loans I and II." Report 15794. Operations Evaluation Department, Washington, D.C. Processed.

The Potential and Limitations of Self-Targeted Food Subsidies

Harold Alderman • Kathy Lindert

Can self-selection of subsidized commodities be used as a mechanism to transfer income to the poor? Evidence from two self-targeting programs, one in South Africa and one in Tunisia, shows that although self-targeting can clearly improve the distribution of food subsidies to the poorest members of society, its power to alleviate poverty and reduce income disparities is limited by preference patterns, income inequality, and the size of the individual subsidies. Self-targeting through quality and product differentiation can be a useful means to reform existing universal subsidy schemes, but it should be considered a transitional tool while the capacity for implementing more precise mechanisms is developed.

Self-targeting—subsidizing only those commodities the target group has indicated it prefers—has been put in place as a way to reform generalized food subsidies and limit participation in these programs to the intended beneficiaries. By choosing to subsidize commodities consumed primarily by the poor, governments have found that they can improve the targeting—and reduce the costs—of food transfers. Self-targeting is attractive because it sidesteps the difficulty of determining how much income people have. By targeting the behavior of the poor, which they reveal, policymakers can avoid having to screen individuals on the basis of their incomes, which they generally do not divulge.

One of the first examples of self-targeting reported in the literature was an experiment in Bangladesh in which sorghum flour was used as an alternative to wheat in ration shops (Karim, Majid, and Levinson 1980). This successful pilot program proved to be better targeted to the poor population than wheat flour (which was itself self targeted) because rice is the preferred grain in that country (Ahmed 1988).

Self-targeting can be also be achieved by quality differentiation. Tunisia subsidizes foods in several product groups (cereals, cooking oil, sugar, milk) according to the consumption preferences of the rich and the poor for various items in each group (Lindert 1995b; Tuck and Lindert 1996). Both Morocco and Egypt have reduced the

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 213–29

costs of their extensive subsidy programs by shifting subsidies toward a dark, rough flour that is consumed disproportionately by the poor and shunned by the rich (Linder 1995a; Ali and Adams 1996). Similarly, yellow maize has been distributed in drought and war relief efforts in Sub-Saharan countries such as Mozambique, where white maize is the preferred staple of the rich (Dorosh, del Ninno, and Sahn 1996).

Although these experiences have been examined in case studies, the literature does not take a broad view of the experience with self-targeting. This article looks at the conceptual, empirical, and practical limitations of self-targeting to determine whether such programs successfully transfer income to the poor and whether they improve the distribution of food subsidies to the intended beneficiaries.

The Self-Targeting Mechanism

Targeted transfer programs that benefit the poor are a core element of the World Bank's three-pronged strategy to reduce poverty, along with broad-based economic growth and human capital development (World Bank 1990). Such transfers aim to reach a target population among the chronic poor who remain outside the economic growth process and to protect the vulnerable during periods of economic change. The rationale for targeting these transfers is based on the premise that, whereas the costs of safety-net programs are essentially the same for all beneficiaries, the social returns for a given level of benefits are higher for the poor than for the wealth[y]. Targeting can improve program efficiency and save resources by concentrating expenditures on those who need them the most. It often entails other costs, however, including the administrative costs of screening potential beneficiaries to identify those who qualify (particularly for means-tested transfers), possible economic losses due to disincentive effects, and the potential loss of political support from those who are excluded because they are better-off (Grosh 1994).

Targeting is generally implemented through three types of mechanisms: an individual income (or needs) assessment, geographic and other group indicators, and self-selection. Self-selection, which has been touted as an administratively simple way to direct transfers to those in need, occurs when the benefits of a transfer scheme are ostensibly made available to all consumers, but the program is specifically designed so that the nonpoor elect not to take up these transfers.

Self-targeting is commonly used in food transfer programs, in which product and quality differentiation discourage consumption of the subsidized product by wealthier consumers. Most consumption-based transfers inherently involve a rudimentary degree of self-targeting, because food represents a larger share of the total expenditure of the poor than of the rich. But generalized food subsidies still typically transfer higher *absolute* benefits to the rich, who purchase larger overall quantities of food. Food transfer programs can be better targeted by selecting certain foodstuffs to carry

higher subsidies than others. Subsidizing foods that are unattractive to the nontarget group can help prevent consumers who are better-off from trying to capture the benefits (Nichols and Zeckhauser 1982). Those commodities that are consumed less as income rises are excellent candidates for such subsidies.

Public employment programs also rely on self-targeting by establishing time requirements and wage rates that make participation unattractive to higher-income individuals (Ravallion 1991). Such programs, however, are based on a different criterion for self-selection. In the case of subsidized goods, the price of a commodity is generally the same for all consumers, but the demand differs according to income. Self-selection for public works programs does not necessarily require a different demand for leisure among the well-off; wages offer the potential for screening because the price of leisure for this group (their wage opportunity) will generally be higher.

Advantages and Disadvantages

Self-targeting has several advantages. First, the information needed to implement self targeted plans is relatively less cumbersome than that required by other targeting mechanisms. The choice of which commodities to subsidize is generally based on survey data on household behavior (consumption patterns), which is less costly to collect than assessments of individual income or income proxies, both of which are subject to substantial inaccuracies. As a result self-selection avoids the problem of asymmetric information regarding income levels (Besley and Kanbur 1988; Besley and Coate 1991; Lindert 1995b)

Second, self-selection mechanisms have the flexibility to respond to changing economic conditions. Although indicator targeting rarely responds to idiosyncratic (individual) fluctuations in income, individuals can shift their commodity choice rapidly in response to changing circumstances. Similarly, more people may choose a subsidized commodity when prices of higher-quality grades of that food rise. In Bangladesh, for example, purchases of subsidized food from the ration shop were sensitive to the price of rice on the open market (Montgomery 1985). This phenomenon indicates a substantial cross price response that can enhance the advantage of self targeting lower grades of a commodity. Because the poor are more sensitive than the rich to changes in food prices, self-targeting is enhanced when prices of nonsubsidized food rise (Timmer and Alderman 1979). Self-targeted programs also may be more easily phased out as incomes climb because beneficiaries voluntarily opt out of the program once they can afford higher-quality foods. This appears to be the case in Pakistan, where demand for rationed flour (which was of an inferior quality) decreased over time, thereby improving targeting (Alderman 1988).

Third, self-targeting may be less vulnerable to bureaucratic corruption and manipulation than other targeting mechanisms (such as means-testing, which requires maintaining a beneficiary roster). Fourth, self-targeting may be less divisive and more

politically acceptable than individual assessment and geographic targeting mechanisms because the decision to participate is made by the individual rather than by the bureaucracy.[1]

Finally, governments also may opt to subsidize commodities that have a higher nutritional content than commodities that are close substitutes. For example, a subsidy on whole wheat flour may shift consumption from more refined flours, with benefits for the long-term health of the population that are in addition to the poverty alleviation objectives of the subsidy.

Self-targeting is not without disadvantages, however. One important drawback to all targeting mechanisms is that some of the very poor may be screened out of the program along with the nonpoor (Subbarao and others 1997). For example, subsidized food products may not be available in poor remote rural areas where subsistence and barter replace cash-based market purchases of food. Experience has also shown that targeting can be imprecise, resulting in large leakages to the nonpoor. This leakage might increase over time if tastes shift toward previously unfamiliar subsidized foods (to subsidized yellow maize, say, in countries where white maize is the norm, or wheat where diets are based on rice). In addition, when lower-quality varieties of commonly consumed food are subsidized, such as high-extraction flour (with some but not all of the bran removed), it can be difficult to disentangle the subsidies on these items from other products made from similar raw materials. This is the case in Morocco, where the subsidy on a high-extraction-rate flour that is well targeted to the poor is injected at upstream levels (meaning closest to the source) of the marketing chain to simplify the payments process (fewer agents and transactions). This self-targeted subsidy has suffered leakages because a fine, white flour preferred by the rich (*farine de luxe*) is made from the same raw material as the targeted commodity (Lindert 1995a).

On occasion, self-targeting is achieved by making the *process* by which a commodity is acquired the means for self-selection. For example, subsidized goods may be available only to those individuals willing to wait in lines or be stigmatized as poor (as when items are available only in state ration shops). This is somewhat different from targeting a commodity that is unlikely to be selected by relatively well-off consumers. In addition, the administrative and other costs of these types of programs result in a deadweight loss to the economy because there is no gain in welfare from this use of resources (Alderman 1987; Ranney and Kushman 1987).

Inherent Limitations to Self-Targeting: Inequality and Demand Parameters

The effectiveness of self-targeting depends on the distribution of income as well as on the commodities being subsidized. A stylized demonstration of these effects on

country with high inequality and one with low inequality with illustrative elasticities representing different commodities is shown in table 1.[2]

Elasticity Effects

The share of government subsidy expenditures that goes to the poor clearly improves with lower expenditure elasticities, regardless of income distribution (elasticity refers to the change in consumption as a result of a change in income). Table 1 shows that substantial equity gains arise from simply reallocating subsidies within the range of positive expenditure elasticities— such as shifting from a commodity with an elasticity of 1.0 to a good with an elasticity of 0.3 (in both the high- and low-inequality countries). The less demand for a good increases with income, the greater the share of subsidies that will go to the poor. Thus, shifting subsidies from commodities such as dairy products or meat (generally with high income elasticities) to commodities like sugar or grains (generally with lower elasticities) could result in greater targeting of subsidies on the poor. Further gains in targeting transfers to the poor could be achieved by identifying commodities with *negative* expenditure elasticities, as shown in table 1.

Table 1 also demonstrates the limits on the *redistributive power* of self-targeted food subsidies. Over a plausible range of income elasticities, it is unlikely that the share of a subsidy going to the poorest two quintiles could be increased much beyond two-thirds of the total transfers (or that accruing to the top three deciles being reduced to less than one-third of the total) by self-selection alone. One illustration of

Table 1. *Determinants of Self-Targeting of Food Subsidies: Stylized Inequality and Demand Parameters*

	Expenditure elasticity					
[illegible]	*1.0*	*0.3*	*0.1*	*0*	*–0.2*	*–0.3*
High inequality						
Share to poorest 10%	2.1	9.5	13.1	15.3	17.5	22.4
Share to poorest 20%	5.4	19.8	25.8	29.0	32.4	39.2
Share to poorest 40%	20.8	40.1	48.7	52.8	56.9	64.5
Low inequality						
Share to poorest 10%	5.7	10.8	12.7	13.5	14.9	17.2
Share to poorest 20%	12.8	21.4	24.3	25.2	27.5	30.9
Share to poorest 40%	30.1	42.9	46.8	48.8	50.6	54.6

Note: Authors' calculations. Decile rankings are by households. The high-inequality example is based on the 1993 South Africa Living Standards and Development Survey (with a Gini Index of 58.22). The low-inequality example is based on the 1996 Albania Employment and Welfare Survey (with a Gini Index of [illegible]).

plausible ranges of income elasticities for subsidized gradation of commodities comes from Pakistan. In that country a subsidy on an inferior quality of flour with an elasticity of −0.13 achieved real—but modest—fiscal savings over the amount that would have been spent if all wheat flour—with an elasticity of 0.34—was subsidized (Alderman 1988).

Moreover, even when inferior goods exist, the budget shares associated with these items limits the extent to which income is transferred to the poor: these products are often consumed in such a small quantity even by the poor that the implicit income transfer embodied in the subsidy is negligible.

Inequality Effects

Although income elasticities indicate how much the purchase of a commodity increases or declines with a change of income, they are not sufficient to indicate the potential of commodity targeting within a country because they do not indicate the degree to which incomes differ. In general, the share of a self-targeted subsidy that goes to the poor tends to increase with higher degrees of inequality. The larger the percentage difference between the incomes of the poor and the nonpoor, the greater the difference in consumption implied by a given negative income elasticity. In the high-inequality country in table 1, shifting subsidies from a luxury good to an inferior good with an elasticity of −0.3 increases the share of the transfer accruing to the poorest by a factor of 10. The difference is only three fold in the low-inequality country. In both examples, the impact of commodity choice is proportionally lower when the target population is a larger share of the total. The disparity across rows in table 1 is greater for all three target groups where inequality is greater. Hence, the lower the inequality, the lower the scope for self targeting of commodity subsidies.

Case Studies: Empirical Impact and Limitations

Two applications of self-targeting, one in South Africa and one in Tunisia, indicate the potential and limitations of the mechanism to transfer income to the poor. Both programs are fairly well-targeted relative to other examples of self selection found in the literature (Alderman 1991; Grosh 1994).

South Africa's Experience with Self-Targeted VAT Exemptions for Food Items

A prevalent school of thought on fiscal policy suggests that value added taxes (VAT) should be based on efficiency criteria alone. This view holds that equity concern can be better addressed with targeted income transfers and similar measures. The

VAT, however, often reflects the government's distributional and fiscal objectives, and for this reason is adjusted to reduce the relative burden on low-income consumers.

The introduction of the VAT in South Africa in 1991 straddled these two viewpoints. Because it was clear that the tax burden would affect the ability of the poorest groups to afford an adequate diet, the government launched a safety net initiative, the National Nutrition and Social Development Programme, aimed at distributing 400 million rand (more than $100 million) annually in community-based food security projects. In addition, to keep their costs down, maize and brown bread were exempted from the VAT shortly after its introduction. By mid-1993, 19 food commodities were exempt, and roughly the same number of additional exemptions had been proposed, including several "luxury" foods, such as meat and dairy products (Alderman and del Ninno 1997).

The fiscal impact of these exemptions is uneven (table 2). The revenue loss from the exemption on maize is similar to the revenues forgone from the exemption on fresh milk and to the combined revenue loss associated with exemptions on brown bread and white bread (for which an exemption has been proposed). The revenue that would be lost from the proposed exemption on meat, however, is roughly equivalent to that for the three other commodities taken together.

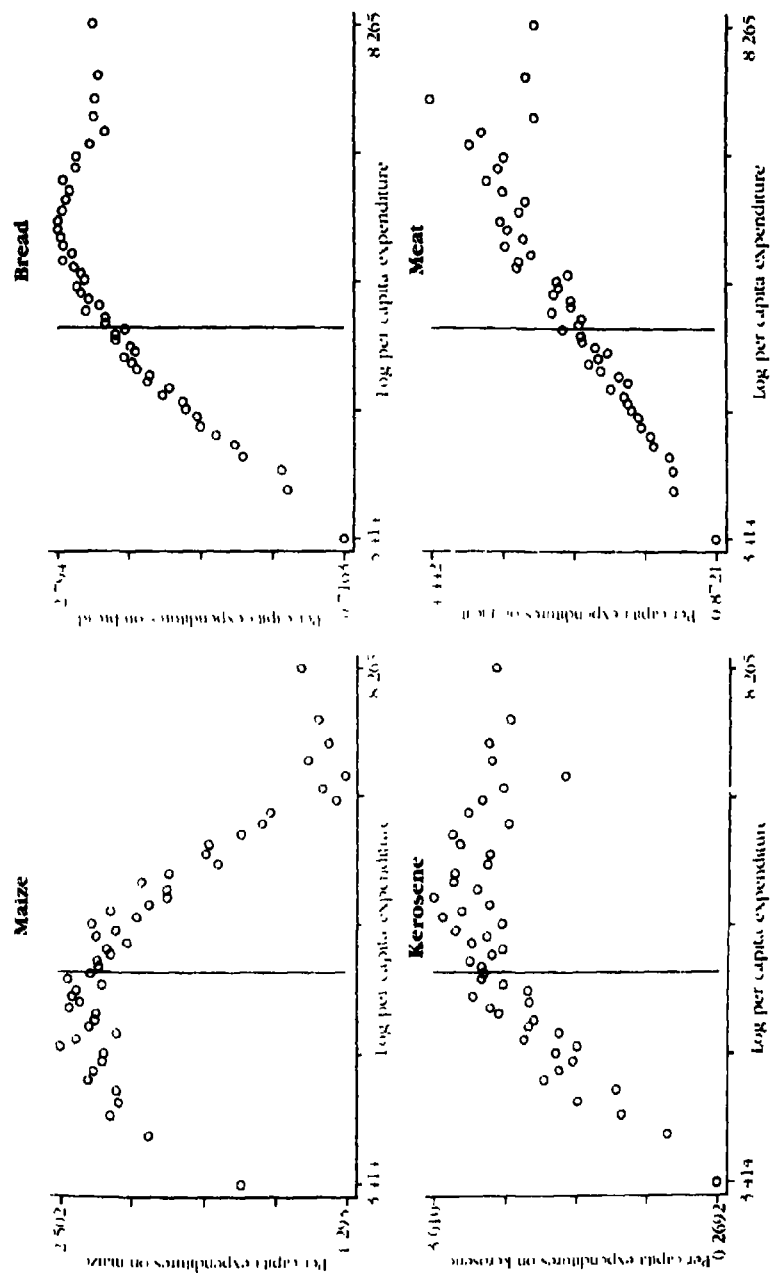
REDISTRIBUTIVE POWER OF SELF-TARGETED TAX EXEMPTIONS. The analysis of the distributional impact of VAT exemptions is conceptually similar to the study of the impact of price subsidies. Standard methodologies have been devised to indicate the efficiency of such taxes (in terms of minimizing economic distortions for a given amount of revenue) as well as the equity impacts (Ahmad and Stern 1991; Deaton and Grimard 1992). Applying this methodology to the 1993 South Africa Living Standards and Development Survey (LSDS) shows that the effects of VAT exemptions differ appreciably across commodities (see table 2).[5]

Maize and kerosene are clearly the preferred commodities for self-targeted price subsidies (or VAT exemptions). The LSDS defines the poor as the poorest 40 percent of households, or 52.8 percent of the population. Thus, only the existing exemption on maize and the proposed exemption on kerosene deliver a share of tax relief to the poor that is greater than their share in the population.[6] The products with the highest leakage of benefits to the nonpoor are meat, milk, vegetable oil, and other dairy products.

Figure 1 confirms that the highest income groups consume less maize than poorer people. Consumption actually increases up to the poverty line, and then declines. This curvature, however, is not captured in the demand estimation employed nor by many common alternative functional forms. Because the figure only shows consumption by households that use kerosene, it masks the fact that the poor consume most of that fuel. As incomes rise, households tend to shift to other sources for fuel

Table 2. *Distribution of VAT Exemptions in South Africa, 1993*

Commodity	*VAT status*	*Fiscal revenue loss (millions of rand)*	*Income elasticity*			*Budget share*	*Value of savings for poor (millions of rand)*	*Percentage of transfers to poor*
			Rural	*Urban*	*Pooled*			
Maize	Exempt	666.9	0.31	0.00	0.02	6.91	436.1	65.4
Fresh milk	Exempt	621.9	1.23	0.78	1.01	2.54	95.3	15.3
Beans	Exempt	102.3	0.90	0.47	0.70	0.78	43.4	42.4
Vegetable oil	Exempt	179.7	0.63	0.45	0.55	1.20	64.5	35.9
Meat	Under consideration	1,807.0	1.29	0.84	1.09	8.27	325.2	18.0
Bread	Brown = exempt							
(brown, white)	White = under consideration	664.1	0.99	0.50	0.80	4.32	221.8	33.4
Other dairy	Under consideration	444.7	1.23	0.78	1.01	2.01	82.2	18.5
Sugar	Under consideration	260.1	0.44	0.29	0.28	2.42	128.0	49.2
Kerosene	Under consideration	152.4	0.31	-0.40	0.11	1.77	100.5	66.0

Source: Alderman and del Ninno (1997).

for cooking and heating. The bread patterns (which include both white and brow varieties) reflect another common problem with survey data—the inability to disti guish commodities by quality. In this particular case, the approach used by Deato and Grimard (1992), which uses unit prices to distinguish quality, is not applicab because prices were collected mainly at the community level, and these show th prices of brown and white bread overlap.

South Africa might seem a fairly unique illustration because poverty there has very strong race-based dimension. As consumption patterns also differ by race, may appear that the self-targeting parallels race-based targeting—a targeting schem that, although theoretically possible, is often politically unacceptable. Indeed, th constitution of the Republic of South Africa forbids such targeting. Self-targetin however, delivers the majority of the VAT exemptions that accrue to the Africa population to the poor among this group. For example, two-thirds of all maize co sumed by Africans is consumed by those Africans in the poorest 40 percent of th total population.

SIZE OF TRANSFER TO THE POOR. Although the tax exemption on maize (and p sibly kerosene) appears to be well targeted to the poor, this subsidy has a ra limited ability to transfer income to the poor. Taking the 12 percent of the bu for this commodity for the rural population—most of whom are poor—and m tiplying it by the size of the tax (14 percent), one finds that the magnitude of total transfer in terms of a reduction in the cost of living is a mere 1.7 percen total expenditures of the poor. Adding the transfer from the subsidy on kerose which is an even more minor share of the budgets of the poor, only brings figure up to 2 percent. Because there are few, if any, other commodities for wh a tax exemption would even be neutral in distributional incidence (let alone w targeted to the poor), it appears that VAT exemptions can have only a modest pact on poverty.

Clearly, alternative price regimes using self-targeting—for example, eliminat the regressive VAT exemptions on fresh milk and vegetable oil and using the r enues for direct subsidies on maize and kerosene—could boost the share of inco supported by self-targeted subsidies. Still, such subsidies would be too small benefit the recipients as much as other targeted transfers in South Africa. For ample, although the means- and age-tested income transfer embodied in the age pension is targeted as efficiently as the VAT (66.2 percent goes to the poor), size of these benefits dwarfs the magnitude of income support generated by t subsidies (Case and Deaton, forthcoming). Thus, while the judicious choice commodities for VAT exemptions can achieve a reasonable degree of targeting, th interventions are still secondary tools in programs designed to provide a safety for the poor.

Tunisia's Experience with Self-Targeted Food Subsidies

In Tunisia the government moved to a self-targeting framework when severe political and administrative constraints frustrated its efforts to cut existing food subsidies (Lindert 1995b; Tuck and Lindert 1996). In the face of a structural adjustment program in the mid-1980s, fiscal pressures dictated cuts in the extensive generalized food subsidy program that had served for several decades as the primary vehicle for transferring income to the poor.[5] The scheme had indeed become quite costly: by the mid-1980s subsidy outlays had ballooned to more than 4 percent of gross domestic product (GDP) and to 10 percent of total government spending (table 3). It was also inefficient: in 1985 the value of the benefits to rich households was twice that of the benefits transferred to poor households.

Initial attempts to reduce the program in the mid-1980s were met with riots, which forced the government to reinstate the subsidies. Policymakers attempted to develop alternative direct transfer schemes to be targeted using individual assessment

Table 3. *Self-Targeted Food Subsidies in Tunisia*

(percent unless otherwise specified)

Year	Product program description	Transfers: To poorest percent	Transfers: To richest percent	Value of transfer to poor as a percentage of total expenditures	Total transfer cost as share of GDP
1985	Prereform universal subsidy program	8[illegible]	1[illegible]	15	4
1990	Prereform universal subsidy program	1[illegible]	20	9	3
1993	Self-targeted reform program				
	Subsidies on self-targeted products only[d]	25	14	6	1
	All food subsidies	21	18	8	2
1993	Simulated results of further self-targeting reforms[f]	2[illegible]	13	4	0.8

a. Data from 1985 Institut National de la Statistique (INS) Household Expenditure Survey, poorest expenditure group = 13 percent of population, richest expenditure group = 12 percent of population.

b. Data from 1990 INS Household Expenditure Survey.

c. Data from 1993 INS Household Expenditure Survey.

d. Semolina, large loaf bread, bulk generic cooking oil, pasteurized reconstituted milk in less-convenient packages.

e. Self-targeted products listed above plus a number of other items.

f. Reducing or eliminating subsidies on poorly targeted items that remained subsidized under the 1993 program using elasticities estimated from an almost ideal demand system.

Source: Lindert (1995b), Tuck and Lindert (1996).

mechanisms in hopes that these could eventually replace the generalized subsidy scheme.[6] High administrative costs, information constraints, and implementation difficulties plagued these programs, however, and leakages to the nonpoor were substantial. Means-tested food stamps were also rejected because of these same administrative constraints and on the grounds that they would be politically unacceptable because they might conjure up images of "wartime" ration cards. The government also explored the possibility of targeting transfers geographically, but this option was abandoned because distinctions between neighborhoods were too obscure in most areas to make geographic targeting effective.

Political, economic, and administrative constraints favored reforms that could be carried out within the existing framework of food price subsidies rather than those that would require an entirely new institutional structure. These concerns led the government to self-target the subsidies to the poor using quality differentiation. One aspect of these reforms involved liberalizing markets to allow private operators market higher-quality food products to attract wealthier consumers. The government also shifted subsidies to narrowly defined items within a product line that perceived by consumers to be of a lower quality because they possess certain unattractive features in their packaging or ingredients. Although the intrinsic nutritional value of these products has been preserved, the perceived "inferior" characteristics these items discourage their consumption by wealthier households.

Self-targeting through quality differentiation was applied to all major subsidized food categories. Subsidies were shifted to semolina, which is disproportionately consumed by the poor, and reduced for pasta and couscous, which do not display so targeted consumption patterns. Subsidies on baguettes, which were consumed virtually exclusively by the rich, were eliminated, while those on a better targeted large loaf of bread were maintained. The bread market was also liberalized to allow private bakers to produce high-quality bread to attract the wealthy customers. Cooking oil subsidies were applied to a generic product purchased from bulk oil drums. Subsidies were 40 percent higher on less-refined brown sugar than on refined white sugar, but all consumers —rich and poor—rejected the brown sugar because they perceive it as "dirty." Finally, milk subsidies were shifted to reconstituted milk packaged in less convenient half-liter cartons, making it less desirable to the rich, who tend to purchase local fresh milk in bottles and in tetrabriks, a type of carton designed for long storage life.

Tunisia's self-targeting efforts have proved to be both effective and politically feasible. Reforms have been economically sound: outlays on the subsidy program were cut to 2 percent of GDP in 1993 and 6 percent of total government expenditures (table 3). Moreover, simulations suggest that these reforms would have reduced subsidy outlays even further—to 1.9 percent of GDP and 5.1 percent of public expenditures—if all other factors including GDP, total government expenditures, and population size had remained at 1990 levels.

REDISTRIBUTIVE POWER OF SELF-TARGETED SUBSIDIES. Reforms have also been effective in improving the distribution of the subsidies. Although demand elasticity analysis failed to uncover true "inferior" goods in Tunisian consumption patterns, the rich did not switch to the subsidized products when subsidies were eliminated on higher quality items.[7] Rather, they substituted previously unavailable unsubsidized luxury varieties when they were introduced as part of the reform program. As a result, reforms shifted the subsidy scheme from one that transferred more absolute benefits to the rich than to the poor to one in which the reverse was true. By 1993 the poor benefited 1.2 times more from the program than did the rich (table 3).

Simulations show that additional self-targeting reforms could further reduce costs and improve the coverage of the program. By eliminating all remaining subsidies on products that are not consumed disproportionately by the poor, the cost of the program could be further reduced to 0.8 percent of GDP and to just over 2 percent of total government expenditures (holding all other factors constant). Assuming that the rich would not shift consumption to the remaining subsidized products, the poor would benefit over two times more than the rich. Self-selection through the choice of commodity is thus an effective instrument to improve the target efficiency of universal subsidies.

It is important to note, however, that the efficiency of self-targeted subsidies in Tunisia, while better than the universal program, is not as sharp as that seen in other transfer schemes. A means-tested food stamp program in Jamaica transferred roughly 15 times more benefits to the poorest quintile than to the richest, as did a geographically targeted food supplementation scheme in Peru (Grosh 1994).

SIZE OF TRANSFER TO THE POOR. The amount of income transferred to poor Tunisians through subsidized commodities is limited, as in South Africa, because total purchases of these products are small and the unit subsidies are slight. In Tunisia subsidies on all food products accounted for 8 percent of the budgets of the poor in 1993, of which those specifically targeted to poor consumers accounted for 6 percentage points. In contrast, the VAT exemptions on maize and kerosene in South Africa amount to less than 2 percent of the budgets of poor households.[8] The list of subsidized products, however, still includes large-loaf bread, more of which is consumed by the upper-middle quintiles (third and fourth) than by the poorest 40 percent (although the poorest 20 percent does consume 20 percent more than the richest). If one considers only the transfers from commodities in which the poor consumed more than the average, that is, those that were truly self-targeted to the poor, the program accounted for only 4 percent of the budgets of the poor in 1993. The Tunisia example, therefore, illustrates some of the tradeoffs between finding commodities that are targeted mainly to the poor and finding vehicles to deliver an income transfer.

Conclusions

How effective is self-targeting in reducing income disparities and alleviating poverty? Self-targeting is clearly preferable to indiscriminate universal food subsidies. A stylized example of demand parameters and income distributions reveals that self-targeting is more effective when subsidies are focused on products with low or negative expenditure elasticities in countries with higher inequality. In South Africa, for example, a careful selection of the products to be subsidized or exempted from the VAT would clearly improve the efficiency of these transfers. In Tunisia as well, self-targeting reforms had the intended results of reducing the exorbitant costs of universal food subsidies and improving distribution to the poor.

Tunisia's experience is particularly germane in that self-selection was central to the subsidy reform, but similar savings have been noted in a number of countries. For example, Egypt reduced food subsidies from 19.5 percent of total government expenditures in 1981–82 to 5.3 percent in 1993–94. Although this reduction was fostered by a 10 percent decrease in the eligible population and a partial convergence of domestic and international prices, most of the cost savings were achie[ved] by focusing on commodities that had negative or negligible income elasticities a[nd] by eliminating subsidies on items with significantly higher elasticities (Ali a[nd] Adams 1996). Thus, subsidies on sugar and coarse flour—and the bread m[ade] from it—which together accounted for slightly more than half of the expendit[ure] on subsidies in 1981–82, made up more than 80 percent of the subsidy bill [in] 1993–94. In Morocco quality differentiation measures, which involved shift[ing] flour subsidies to a coarser grade, reduced the cost of flour subsidies from 1.8 p[er]cent of GDP in the mid-1980s to 0.34 percent by 1995 and greatly improved c[ov]erage of the targeted population.

Nevertheless, the effectiveness of self-targeted commodity subsidies is limit[ed]. Even under "optimal" circumstances (high inequality and low or negative exp[en]diture elasticities), at least one-third of total transfers can be expected to leak to t[he] nonpoor. This result was observed in the case of self-targeted maize subsidies [in] South Africa and Tunisia. Moreover, self-targeted food subsidies have only a li[m]ited impact on poverty alleviation. Our study is consistent with Sah's (1983) analy[sis] of the limits of redistribution through indirect taxes. Although this investigat[ion] does not presume that all revenues are raised by indirect taxes, the amount [of] income transferred to the poor is constrained by the importance of subsidiz[ed] products as well as by the size of the subsidies. Self-targeting can play a useful r[ole] in reforming *existing* universal food subsidy programs, until governments deve[lop] the administrative capacity to effectively deliver means-tested direct transfer pr[o]grams to alleviate poverty.

Notes

Harold Alderman is principal economist in the World Bank's Development Economics Research Group. Kathy Lindert is an economist in the Human Development Department of the World Bank's Latin America and Caribbean Regional Office.

1. This is not inherent. There are political costs as well as advantages to subsidizing inferior goods. This may explain why many pilot self-targeted programs have had short lives. The Bangladesh sorghum experiment was discontinued for a variety of reasons after a year of operation. Similarly, Pakistan replaced its self-targeted ration shops with a generalized price subsidy. In Tunisia, the self-targeting of lower-quality school notebooks was abandoned because the stigma associated with visibly separating the rich from the poor, particularly for a *durable* item consumed by children, was deemed politically unacceptable shortly after the notebooks were introduced.

2. This illustration uses data from South Africa, with a Gini Index of 58.22, to illustrate a high-inequality country, and Albania, with a Gini Index of 27.61, as an example of a low-inequality country. The examples are stylized, with per capita consumption determined by a constant elasticity with respect to per capita expenditure. When this elasticity is zero, per capita consumption is constant. Nevertheless, the poorest deciles of *households* (ranked in terms of expenditure per adult equivalents) receive more than 10 percent of the subsidy since these deciles have more individuals. Clearly, ranking by decile of *individuals* would give a neutral pattern.

3. The data set is available to the public on the following site: www.worldbank.org/html/prdph/lsms/country/za94/za94home.html#top

4. However, it is noteworthy that this distribution pattern would not be expected using income elasticities generated from an almost ideal demand system. A common rule of thumb suggests that if a commodity has an income elasticity of zero, the share of a subsidy or tax exemption to the poor would be the same as its population share, yet the pooled elasticities for both kerosene and maize are slightly greater than zero. Of course, it is not necessary to estimate an income elasticity to indicate the share of benefits accruing to a population group, although price elasticities are needed for a complete analysis of tax efficiency.

5. The general consumer subsidy program covered a wide range of products, including semolina, couscous, pasta, bread, flour, cooking oil, sugar, and milk. Subsidies were universal and available to anyone who chose to purchase the subsidized products in any quantity desired. The cost of the subsidies was borne entirely by the government (taxpayers), as the transfers covered the gap between producer prices and the artificially low consumer prices.

6. Direct Assistance schemes include the Needy Families Program (cash transfers) and the *Union Tunisienne de Solidarite Sociale*, which is responsible for low-income food rations and cash transfers for the elderly and handicapped (World Bank 1993).

7. Expenditure elasticities for the well targeted items estimated using an Almost Ideal Demand System, while not negative, were low, ranging from 0.17 to 0.37 (Lindert 1995b).

8. These benefits are comparable to Tunisia's main other income transfer program at the time, the Needy Families Program, which suffers from the traditional difficulties associated with indicator targeting: complex administration, failure to update eligibility lists, lack of flexibility, and substantial errors of inclusion and exclusion (Tuck and Lindert 1996).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Ali, S. M., and Adams, Richard. 1996. "The Egyptian Food Subsidy System: Operations anc on Income Distribution." *World Development* 24(11):1777–95.

Ahmad, Ehtisham, and Nicholas Stern. 1991. *The Theory and Practice of Tax Reform for De Countries.* Cambridge: Cambridge University Press.

Ahmed, Raisuddin. 1988. "Structure, Costs, and Benefits of Food Subsidies in Bangladesh.' Pinstrup-Andersen, ed., *Food Subsidies in Developing Countries.* Baltimore, Md.: Johns H University Press.

Alderman, Harold. 1987. "Allocation of Goods through Non-price Mechanisms: Evidence tribution by Willingness to Wait." *Journal of Development Economics* 25:105–24.

———. 1988. "The Twilight of Flour Rationing in Pakistan." *Food Policy* 13(3):245–56.

———. 1991. "Food Subsidies and the Poor." In George Psacharopoulos, ed., *Essays on Equity, and Growth.* Elmsford, N.Y.: Pergamon Press.

Alderman, Harold, and Carlo del Ninno. 1997. "Poverty Issues for Zero Rating VAT in So rica." World Bank, Development Economics Research Group, Washington, D.C. Proc

Besley, Timothy, and Ravi Kanbur. 1988. "Food Subsidies and Poverty Alleviation." *The E Journal* 98 (September): 701–19.

Besley, Timothy, and Stephen Coate. 1991. "Public Provision of Private Goods and the Red tion of Income." *American Economic Review* 81(4, September):979–84.

Case, Anne, and Angus Deaton. Forthcoming. "Large Cash Transfers to the Elderly in So rica." *Economic Journal* 108.

Deaton, Angus, and Franque Grimard. 1992. *Demand Analysis and Tax Reform in Pakistan* Bank, Living Standards Measurement Study Working Paper 85. Washington, D.C.

Dorosh, Paul, Carlo del Ninno, and David Sahn. 1996. "Market Liberalization and the Food Aid in Mozambique." In David Sahn, ed. *Economic Reform and the Poor in Africa* C Clarendon Press.

Grosh, Margaret. 1994. *Administering Targeted Social Programs in Latin America. From Plat Practice.* Washington, D.C.: The World Bank

Karim, Rezaul, Manjur Majid, and James Levinson. 1980. "The Bangladesh Sorghum Exper *Food Policy* 5(1):61–63

Lindert, Kathy. 1995a. "Food Subsidies in Morocco." World Bank, Human Development I ment of the Latin American and Caribbean Regional Office, Washington, D.C. Process

———. 1995b. "Quality Differentiation and the Self Targeting of Food Subsidies in Dev Countries." Ph.D. Dissertation, Department of Agricultural Economics, University of C nia at Davis. Processed.

Montgomery, Roger. 1985. "Statutory Rationing and Modified Rationing. Causes and E *Bangladesh Development Studies* 13(February):109–16.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrict Recipients." *American Economic Review Papers and Proceedings* 72(2):372–77

Ranney, Christine, and John Kushman. 1987. "Cash Equivalence, Welfare Stigma, and Food St *Southern Economic Journal* 53(4):1011–27

Ravallion, Martin. 1991. "Reaching the Rural Poor through Public Employment: Argumen dence, and Lessons from South Asia." *The World Bank Research Observer* 6(2):153–76

Sah, Raaj Kumar. 1983. "How Much Redistribution Is Possible through Commodity Taxes? *nal of Public Economics* 20(1):89–101.

Subbarao, Kalandhi, and others. 1997. *Safety Net Programs and Poverty Reduction. Lessons fron Country Experience.* Washington, D.C.: The World Bank.

Timmer, C. Peter, and Harold Alderman. 1979. "Estimating Consumption Parameters for Food Policy Analysis." *American Journal of Agricultural Economics* 61(5):982–9.

Tuck, Laura, and Kathy Lindert. 1996. *From Universal Food Subsidies to a Self-Targeted Program: A Case Study in Tunisian Reform.* World Bank Discussion Paper 351. Washington, D.C.

World Bank. 1990. *World Development Report 1990: Poverty.* New York: Oxford University Press.

World Bank. 1993. *Republic of Tunisia: The Social Protection System in Tunisia.* Report 11376-TUN. Washington, D.C.

Targeting Revisited

Dominique van de Walle

Public spending programs aimed at alleviating poverty can either be broadly targeted at categories of spending or narrowly targeted at types of people. Each approach has benefits and costs to the poor. It is often claimed that narrow targeting of the poor will allow governments to reduce poverty more effectively and at lower cost. But narrow targeting often has hidden costs, and once these costs are considered, the most finely targeted policy may not have any more effect on poverty than a broadly targeted one. Both approaches also have hidden benefits, although less is known about their impact. Targeting can help, but it is not a cure-all. Reducing poverty calls for broadly targeted social sector spending combined with narrower targeting of cash and in-kind transfers to specific groups. It is also important for governments to experiment with schemes that offer better incentives, to carefully monitor the costs and outcomes, and to be flexible and pragmatic in their policy responses.

How can public spending assist those who have been left behind by economic growth? It may take a long time for some subgroups in society to catch up. Some, such as the elderly and disabled, may never achieve that goal. Others may well be hurt in the short run by policy reforms that ultimately would benefit the poor. There may be related concerns about regionally unbalanced growth. How can public spending help deal with vulnerability? Incomes can be highly variable over time, particularly in poor rural economies where the population is vulnerable to risks induced by uncertain weather, relative price shifts, or the collapse of community-level support systems during crises. How can public spending meet these objectives adequately with limited resources?

The answer often given to all three questions is "targeting." Those who have been left behind, or who are vulnerable to risk, can (it is argued) be reached most cost effectively by concentrating limited public resources on narrowly defined "target" groups within society.

In practice, there are two approaches to raising the living standards of the poor with public spending. The first involves targeting types of spending and can be called broad targeting. Under this approach no attempt is made to reach the poor directly as indi-

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 231–48

viduals. Instead, gains are achieved by targeting types of spending that are relatively more important to the poor. Spending on basic social services, such as primary education and primary health care, is one example. Rural development is another. The second approach entails targeting categories of people. Under this approach, which can be called narrow targeting, benefits are intended to be targeted directly to the poor. Such policies include food stamp schemes targeted to poor mothers, innovative microcredit schemes aimed at rural landless women, and development programs that focus on poor geographical areas. Within each of these categories are differences in how much the program relies on administrative targeting and how much it depends on self-selection based on behavioral responses to the incentives built into the program. Each approach has benefits and costs to the poor and to others.

This article reviews the case for targeting and the relevant recent evidence for developing countries. Much has already been written on this topic. A good starting point is Besley and Kanbur (1993), who cautioned against taking too simplistic a view of targeting and identified the potential costs associated with such efforts. Almost a decade later, it is still common for empirical work and policy discussions to disregard the fact that targeting is almost never costless. But research has progressed on these issues, and the results from careful studies offer some guidance on cost-benefit assessments.

The article revisits the issues surrounding targeted programs in light of the new research and also points to continuing deficiencies in our knowledge. Currently available data and research methods often make it difficult to assess the full costs and benefits of targeting, which are often hidden. Costs and benefits are frequently not properly accounted for in policy assessments and hence are ignored in policy decisions. Researchers are getting better at measuring some factors, such as how utilization of targeted services varies by income group, but other aspects, such as the behavioral responses of participants, administrators, and governments, are still being neglected. People routinely underestimate how difficult it is to target the poor. This point has important implications not only for policies directly aimed at poverty alleviation, but also in areas such as macroeconomic policy. For example, it is often argued that targeted interventions can be used to compensate losers from macroeconomic reforms, yet little consideration is given to how this is to be done. Once difficulties in practice are acknowledged, it is easy to see that more attention needs to be paid to the timing and sequencing of macroeconomic operations to take into account the government's inability to compensate losers in the short run.

Broad Targeting

Although narrow targeting has received the bulk of attention and is the main focus of this paper, the allocation of budgets among categories of public spending can also

be a means of targeting the poor. By looking first at broad targeting, one can also better understand some of the motivation for narrow targeting.

The basic principle underlying broad targeting is that some categories of public spending matter more to the poor than do other categories. For example, spending on basic social services is found to often benefit the poor. Money spent on primary education, for example, is likely to reach more poor children than money spent on secondary or tertiary education, because many poor students will have dropped out of the higher levels, unable or unwilling to afford the opportunity and other costs of continued schooling. Figures 1 and 2 show the distribution across expenditure groups of public spending for health and education in Indonesia, and for education in Côte d'Ivoire. Individuals are ranked by household per capita expenditures and divided into population quintiles or deciles. The utilization of sector services—school attendance at different levels or visits to types of health care facilities—are tabulated for each group and multiplied by the unit costs of providing each specific service to get the total spending implicitly directed to each group. These amounts are expressed on a per capita basis and plotted against the expenditure groups.

In a pattern that is frequently repeated across developing countries, spending on primary education tends to favor the poor, with per capita amounts declining as living standards rise, while spending on higher education, especially tertiary education, benefits primarily the richest group. Similarly, in the health sector, hospital expenditures confer the highest transfers to the top deciles, while spending on health care centers that dispense basic care is much more neutral across expenditure groups. Such evidence has lent credence to the case for broad targeting of public spending to basic social services.

Several additional factors strengthen the case. Better health and basic education, access to safe water, and basic physical infrastructure raise poor people's well-being and may also raise their productivity and incomes. The nonpoor often already have enough of these services, so the added spending reaches the poor. In Malaysia, for example, Hammer, Nabi, and Cercone (1995) find evidence indicating that between 1985 and 1990, marginal expenditures on safe water supply went primarily to households in Malaysia's poorer states. This happened in part because households in the rich states already benefited from this service. It is often argued that basic services tend to be underprovided by the private sector in the absence of government intervention, especially to the poor. For example, poor parents may not be willing to pay the full costs of their children's schooling, particularly when (as is the case for girls in some societies) parents share little in the future benefits of that schooling. Finally, because it makes no socioeconomic distinctions between who can and who cannot participate, broadly targeted spending is perceived to be more politically popular than narrow targeting, and hence more sustainable.

Figures 1 and 2 make the case for broad targeting of public spending to basic social services. How compelling is that evidence? The figures rely on the "benefit

Figure 1. *Subsidy Per Capita by Decile, Indonesia 1989*

Education

Rupiahs year

1,000

Primary

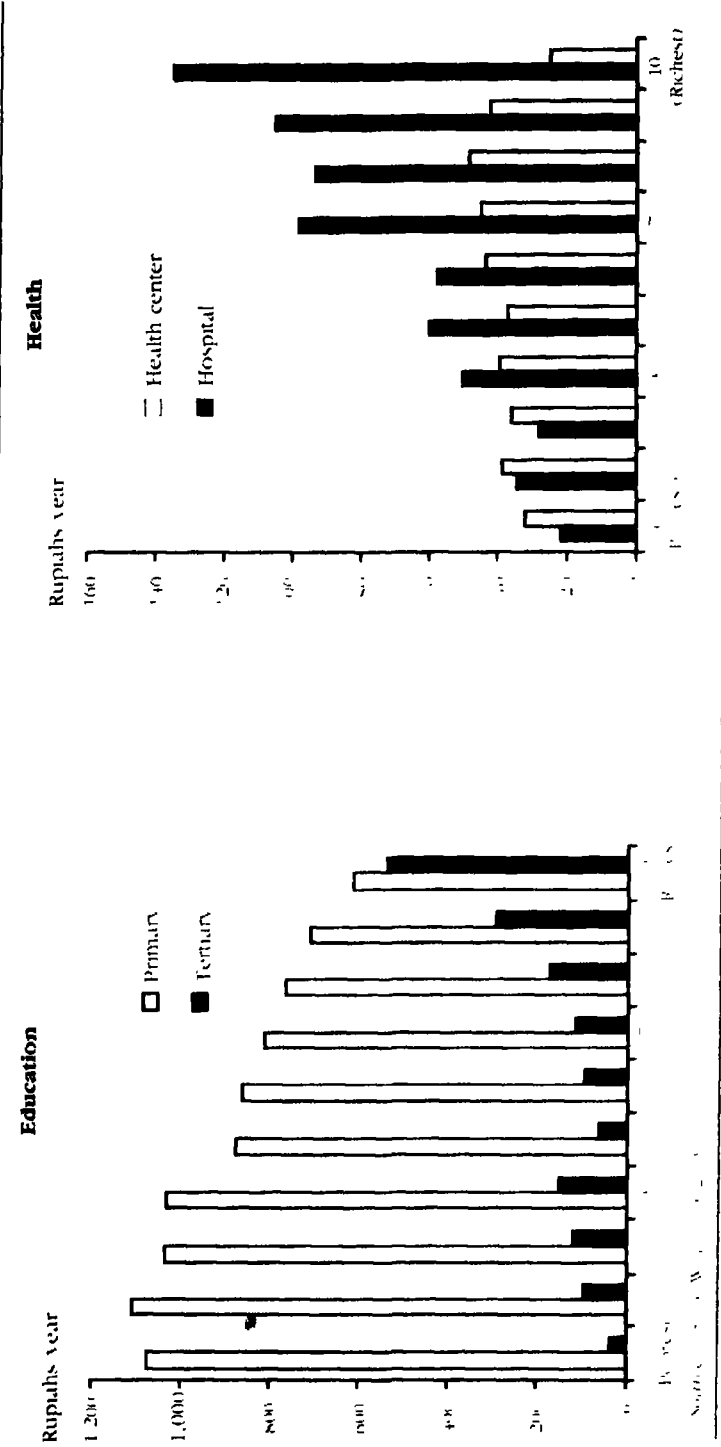
Tertiary

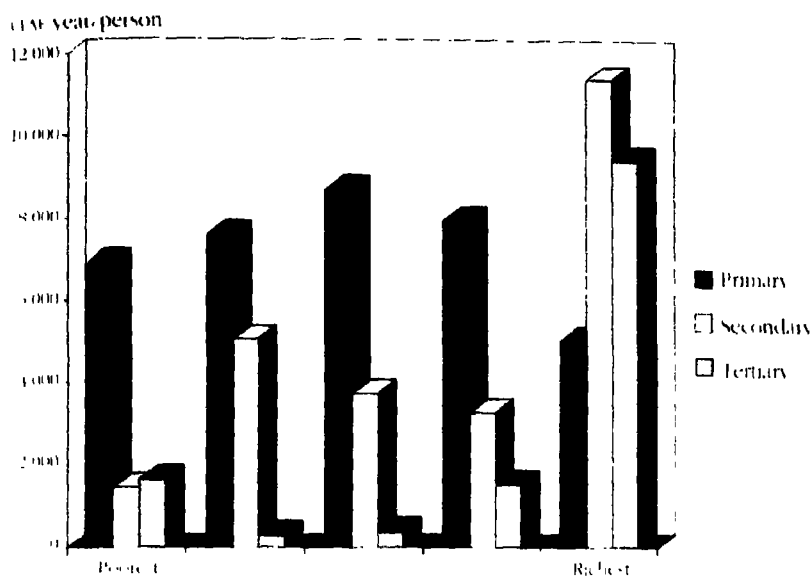
Health

Rupiahs year

Health center

Hospital

10
(Richest)

gure 2. *Education Subsidy Per Capita by Quintile, Côte d'Ivoire, 1995*

Source: [illegible]

incidence" approach to measuring the distributional impact of public spending. Although programs are often analyzed in this way, the method makes a number of strong assumptions that can affect the conclusions drawn and hence the policies advocated (van de Walle 1998). Leaving aside these issues, several other arguments have been made against broad targeting as an instrument for reducing poverty.

The most common criticism is that broad targeting can be expensive. Universal subsidies on basic services also reach many who can afford to pay, as illustrated in figures 1 and 2. Given this "leakage," a broad targeting program can be a costly way to reduce poverty. Even when the incidence of poverty is high, there are likely to be differences in the severity of poverty, and society is likely to put a priority on helping the worst-off first.

Broad targeting may also be a clumsy instrument for reaching some of the poor. Some groups have specific needs that will not be served by the free provision of safe water or basic social services or other categories of spending that are suited to broad targeting. Similarly, broad targeting may be inefficient in meeting specific objectives. A central role of public spending is to protect the vulnerable and provide insurance for downturns and variability in living standards. In a drought or natural disaster the provision of basic social services is unlikely to provide the necessary safety

net. Such concerns have prompted policy discussions to turn to narrow targeting as a means of achieving higher impacts on poverty at lower cost.

Narrow Targeting

Narrow targeting is defined as a deliberate attempt to concentrate benefits on poor people—whatever the type of spending. It has been a particularly popular policy proposal in the context of recent worldwide efforts to reduce budget deficits and public spending while still protecting the poor.

Types of Narrow Targeting

Narrowly targeted schemes are based on one of two principles—or a combination of both. The first is indicator targeting (also called categorical targeting). Such a strategy identifies a characteristic of poor people (an indicator) that is highly correlated with low income but that can be observed more easily and more cheaply than can income (Besley and Kanbur 1993).[1] The indicator is then used as a proxy for income to identify and target poor people. A typical example would be the region of residence; geographical targeting has been popular with many governments and in World Bank projects. Alternatively, landholding class, gender, nutritional status, disability or household size are often used to identify beneficiaries. In the former Soviet Union and East European countries, for example, household size is used to distribute family allowances. Bangladesh's Grameen Bank combines correlates: microcredit is targeted to rural women from landless or near-landless households.

The second approach is self-targeting. Instead of relying on an administrator to choose participants, these schemes aim to have beneficiaries select themselves through creating incentives that will induce the poor and only the poor to participate. Self-targeting works by incorporating a cost of participation into the design of the scheme. This cost should rise as income rises while the benefits remain the same; thus, depending on where the benefit level is set, the nonpoor are effectively screened out. For example, public employment schemes use work requirements to help screen out the nonpoor; subsidy programs support items that the poor consume but the rich do not; and other controls rely on waiting time, stigma, and lower "packaging" quality of goods and services to dissuade usage by the well-off.

Alderman and Lindert discuss self-targeting through food subsidy schemes in detail in this volume. They describe Tunisia's approach to converting a fiscally unsustainable universal food subsidy to a smaller, more pro-poor program through self-targeting (also see Tuck and Lindert 1996). Limited by the absence of any great difference in the consumption baskets of different income groups, the Tunisian government's solution was to differentiate goods not by their intrinsic quality, b

through their outer packaging. "Inferior" goods—perceived by the rich to be of lower quality—were created, and the subsidies were transferred to them. For example, milk was packaged in small, flimsy plastic containers that had to be emptied once opened. Both the size and the inconvenience deterred purchases by the rich. The inconvenience of buying goods in these forms of packaging was the "cost of participation" in this case, and it tended to be higher for the nonpoor. This policy was reinforced by creating unsubsidized superior goods (milk in liter cartons with attractive labeling) and by liberalizing the sale of imports and other high-quality substitutes that appeal to the rich. The policy succeeded in significantly lowering program costs, avoiding political unrest, and concentrating program spending on lower income groups.

The cost of participation promotes self-selection and helps identify who is poor and who is not. It can also be designed to deter the poor from becoming dependent on the scheme (Besley and Coate 1992). The programs are designed to maximize income gains but minimize gains to other aspects of well-being. In this way, they seek to maintain incentives to move out of dependency and poverty. The workhouses for paupers in 19th century England were a classic example of the deterrence argument. The authorities made relief as uncomfortable as possible by requiring beneficiaries to enter the workhouses (Besley, Coate, and Guinnane 1996). One truly had to be desperate. Finally, self-targeted schemes have the added benefit that power is taken away from civil servants and hence incentives toward corruption and favoritism are reduced.

Both types of narrow targeting offer the hope of avoiding two commonly identified errors of targeting: a leakage of benefits to the nonpoor; and imperfect coverage of the poor (Grosh 1995; Cornia and Stewart 1995). Such errors of targeting often influence policy advice. For example, an unfavorable evaluation of the Food for Education program in Bangladesh (World Bank 1997) argued that the benefits reached only 15 percent of pupils and that the leakage was 14 to 26 percent (depending on the target group definition). Others, however, have argued that the low coverage was justified because those who were not covered were unlikely to need any incentive to keep their children in school. Moreover, the leakage to the nonpoor was not high compared to other antipoverty programs (BIDS 1997; Wodon 1998). By these broader criteria, the program looks like a worthy effort. The key point is not whether a scheme avoids errors of targeting, but how well it meets its stated objectives given budget constraints, the information that is available to policymakers, and the behavioral and political responses to targeted interventions.

What Are the Costs of Narrow Targeting?

Narrow targeting, like broad targeting, has costs. Three types of costs are associated with narrow targeting: administrative costs, costs that arise from incentive effects or

behavioral responses, and costs that result from the ramifications of political e
The first is typically (although not always) easily measured, but the second
types are often hidden and require closer scrutiny.

ADMINISTRATIVE COSTS. The first cost arises because policymakers have i
information, and obtaining that information entails costs. Identifying wh
with precision and avoiding leakage by reaching only the poor often ne
high administrative costs. This is a particularly serious problem in developi
tries where most of the poor live in rural areas and policymakers do not have
information on their incomes. Incomes typically vary considerably over
even if reliable data are obtained, the frequent need to reevaluate the inf
can be expensive. Although targeting specific indicators holds the promise
ing administrative costs, even the best indicators tend to be imperfectly c
with low incomes. Combining a few indicators often improves targeting
affect administrative costs as well (Baker and Grosh 1994).

Leakage and administrative costs can eat into a program's budget. A mea
rice subsidy scheme in India's Andhra Pradesh State is illustrative. Radhakri
others (1996) found that for every rupee transferred to a poor person, the
spent 6.4 rupees. Of the additional 5.4 rupees, 3.6 rupees reached the non
1.8 rupees were spent on administration. The potential deadweight costs to
attributable to incentive effects have not been studied, but if they exist, th
further cut into the amount transferred to the poor. The primary problem a
be the substantial transfers to the nonpoor. One can surmise, however, t
sures to avoid leakage might well result in higher administrative costs rat
greater coverage of the poor.

INCENTIVE EFFECTS. Some beneficiaries and nonbeneficiaries may alter the
ior as a result of a policy. Self-targeted schemes often require participants t
their behavior. In the case of indicator targeting, the potential recipients
spending face incentives to misrepresent their incomes and change their
either to become or simply to remain eligible. It can be difficult to find an
that can be identified accurately and at little cost, is strongly correlated with
and cannot easily be manipulated. So, for example, people may change t
dence to gain coverage in a plan that is geographically targeted. Programs d
female-headed households have been blamed for encouraging the break
nuclear family households. Targeting benefits to the unemployed may dam
efforts to find work.

The incentive problem can be severe for those near an income cutoff |
eligibility. Earn a little less and you qualify for a transfer; earn a little more
will be disqualified. Indeed, if the loss of transfer is larger than the income g

there will be no incentive to escape poverty. This "poverty trap" arises when targeting creates a 100 percent marginal tax rate on the poor.

A few studies have attempted to measure the costs associated with behavioral responses to policies aimed at raising the incomes of the poor. One study of Sri Lanka's targeted food stamp program offers an interesting example in examining whether households that received food stamps changed their work effort. The targeted scheme was introduced in 1979 to replace a long-standing general food subsidy. Sahn and Alderman (1995) found that workers in recipient households, both men and women, reduced the number of hours they worked. On average, rural workers cut back almost three days a month, effectively reducing the average monthly net household income gain to 59 rupees, from 91 rupees. This may not be a bad outcome. The recipients' well-being has assuredly increased. Leisure is valued. Policymakers whose objectives include welfare more broadly conceived than income alone will be pleased. But if the policy objective is to maximize income gains, the outcome may disappoint.

Behavioral responses may occur on the part of nonbeneficiaries too. The benefits of public support can be shared between direct recipients and the donors of private transfers. For example, a study of South Africa finds that each rand of public pension support provided to the parents of migrant workers decreased remittances to the parent from the migrant by as much as 0.2 to 0.4 rand (Jensen 1998). This is not considered a deadweight loss as long as positive social value is attached to the implicit gains to migrant children.

Like indicator targeting, self-selection can entail hidden costs arising out of changes in behavior induced by participation costs. The latter costs help focus benefits on the poor, but they also reduce the net transfer benefits, although this fact is often ignored. For example, a review of cost-effectiveness calculations for Bangladesh's targeted food-grain distribution programs concludes that "programs that have work requirements or other obligations (such as sending children to school) have proved to be much more cost effective than food rations" (Subbarao and others 1997:49). The claim is based on the fact that "obligations" greatly improve targeting efficiency and that the underlying cost calculations account only for administrative and food transfer costs. This conclusion could be quite deceptive because the costs to the poor of those obligations have been omitted. Forgone incomes could be non-negligible. The opportunity cost of children's time and other schooling costs might be large. The rationale for programs requiring children to attend school in exchange for transfers is precisely that poor parents cannot afford to send their children to school. The exchange will be acceptable only if the household achieves a net gain. Although the benefit for participants is positive, that gain is likely to be overestimated by conventional cost-effectiveness calculations.

For example, one study evaluates the value of transfer benefits in a Jamaican school-feeding program designed to self-select beneficiaries through the nature of the food

provided (Jacoby 1997). Targeting performance is assessed to be high, but so are the costs. Factoring in behavioral responses as well as administrative costs, it costs 2.60 Jamaican dollars to achieve a gain of 1 Jamaican dollar per child.

POLITICAL ECONOMY. The middle classes are often the primary beneficiaries of public social spending. The poor are largely left out, while the rich have alternatives in the private sector at home or abroad. Spending that is narrowly targeted toward the poor will tend to be associated with a contraction of benefits to the middle class, which is often the government's most vocal and politically important constituency and which may then cease to support poverty reduction efforts. Especially in situations where the prospective beneficiaries of targeted policies are marginalized, it is possible to end up spending less on the poor (Gelbach and Pritchett 1996). In Sri Lanka, for example, the nominal value of the food stamps was never changed between 1979 and 1989, and their real value was halved through inflation. Some have argued that the value of the food stamps was allowed to erode without resistance because beneficiaries of the program lacked political clout, while the more powerful middle class did not support the targeted scheme (Anand and Kanbur 1990).

The U.S. experience also provides evidence for this view and for the often heard adage "programs for the poor are poor programs." The electorate supports cos universal programs such as social security, from which it expects to benefit, but i the targeted means-tested welfare programs such as Aid to Families with Dependi Children—a cash transfer program for poor single mothers and their children

Clearly, many factors influence the political economy of poverty alleviatio programs. Under certain sociopolitical environments, taxpayers may be more o less, inclined to support programs that efficiently reach the deserving poor and save resources for other publicly provided goods (Besley 1997). Information a public relations campaigns have helped to build societal support. A timely awa ness campaign helped secure popular acceptance of Tunisia's plan to reforn universal food subsidy to a more narrowly targeted scheme (Tuck and Lind 1996; Alderman and Lindert in this volume). Polls showed that the campaig which measured the costs of the old program in terms of the roads, hospital be and manufacturing jobs that could otherwise have been secured, successfully co vinced a majority of the population that the change was necessary and desirab Strategies to promote the participation and support of diverse groups appea have helped the sustainability of pro-poor policies within social funds in Boliv and Zambia (Subbarao and others 1997). Design specifics also matter. For e ample, the public is often more sympathetic to food and other in-kind transfe and favors work requirements in exchange for public assistance rather than ca payments with no strings attached. Such burdens to beneficiaries are perceived help weed out the undeserving poor. The macroeconomic environment also ma

s. People who are economically content are more likely to back programs that nefit others. The political environment and its determinants are important fac-s that politicians and policymakers ignore at their peril.

hat Are the Benefits of Narrow Targeting?

mmon methods of assessment may also obscure some of the potential benefits of rrow targeting. Assessments of the benefits from geographical targeting provide an ample. Several studies examine the potential impact on poverty of a predetermined dget optimally allocated across regions. The static gains are often found to be dest, in essence reflecting the fact that the poor are heterogeneous. Not all of em can be identified by the same indicators. For example, even though most coun-es have regions that are poorer than others, not all the poor live there, nor do all rich live elsewhere. Hence, geographic targeting will often benefit some of the h and will bypass and perhaps even tax—some of the poor who live in the better-off areas (Datt and Ravallion 1993; Ravallion 1993).

Recent work, which allows for the potential dynamic effects of programs, suggests that static assessments may considerably underestimate the benefits. Gains may percolate through and strengthen over time through the positive external effects of development in the poor region on the productivity of private investments by the poor. Measuring such effects is difficult and requires data that are often unavailable. In a study assessing the effects over time of development programs that were geographically targeted to poor areas in China, Jalan and Ravallion (1998) found the expected imperfect coverage of the poor and leakage to the nonpoor. But they also found that the programs had a positive impact on growth rates in the targeted areas. In this case, the short term gains to the poor were less than the long-term gains.

Similarly, static assessments have tended to underestimate the benefits of land-contingent targeting (meaning that the beneficiaries are selected solely on the basis of land ownership). Many of the poor in rural areas of South Asia are landless. Yet some poor rural households may have significant amounts of land, and more than some of the rural wealthy. In Bangladesh, Ravallion and Sen (1995) concluded that redistributing income from those with ample land to the landless had only a small effect on poverty even though landholding size is positively correlated with incomes. The authors found that the impact was somewhat higher once the higher productivity of the transfers to small—but often credit-constrained—landholders was considered.

Several other potential hidden benefits are still unresearched, including longer-term benefits of policies that succeed in changing the productive abilities of the poor in increasing their ability to cope with risk and other areas where markets fail; direct benefits from the assets created by public works schemes as well as benefits from potential impacts on employment and wages outside the schemes; and gains

from receiving public support rather than being dependent on rich relative patrons.

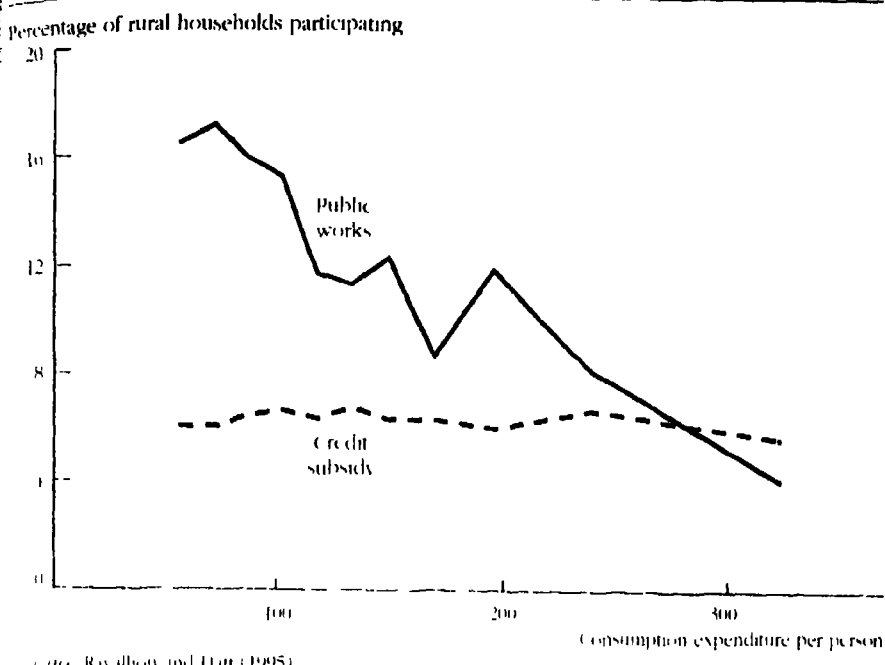
A Case Study of a Relief Work Program

In an effort to stave off the potentially disastrous consequences of a bad drou India's Maharashtra State introduced a public employment scheme in 1973. program's popularity and success in smoothing consumption for the vulnerabl to its permanent deployment as the employment guarantee scheme (EGS). The pays a low wage for unskilled manual work on rural infrastructure projects suc road maintenance, reforestation, and small irrigation works. Anyone who show is given work. The idea is that for only the poor will the cost of participatio primarily that one must forgo other employment or leisure— be sufficiently low them to turn up. But, since few of the able-bodied can afford to be completely this cost will rarely be zero. In assessing impacts on the poor, it is necessary to net this and any other costs to the poor. The potential longer-term benefits that are captured by the wages earned must also be factored in. The final impact will function of the targeting effectiveness of the indirect gains and of the costs associ with targeting: administrative, behavioral, and political.

Evidence from Two Schemes

There has been much debate in India about two poverty reduction schemes Integrated Rural Development Program, a subsidized credit initiative targeted low-income individuals; and public employment schemes (including the empl ment guarantee scheme) that rely on self-targeting. Ideally, one would like to co pare these programs to see which has a greater impact on poverty. Unfortunately data to do so credibly are not available. But the targeting performance of the programs can be compared for Maharashtra State.

Figure 3 plots the percentages of rural households who participated in either sch during 1987–88 against household per capita expenditure levels (Ravallion and 1995). If the schemes were well-targeted, participation should eventually fall to as expenditures rise. It is clear that participation in the employment guarantee sche is highest for the poorest households and tapers off with rising expenditure le But the integrated rural development program data plots more or less a straight participation has no relationship with expenditures. These data suggest that the is much better targeted and that means-testing does not assure better targeting this specific case one can only conjecture about why this was so. Some individua probably understate their incomes to qualify for loans. Another possibility is

Figure 3. *Public Works vs Means-Tested Credit Subsidy, Maharashtra, India, 1987–88*

corrupt program administrators may have wide latitude to choose beneficiaries, help their friends or families, and establish patronage relations

The Costs of Targeting

The EGS is well-targeted; it removes power from administrators, and so is not subject to the same corrupt practices. But this is only one determinant of its impact on poverty. What about the costs of targeting? What benefits accrue to the poor indirectly from public works employment? Let us first consider the political economy of the EGS. Historically, the scheme has enjoyed wide support, in large part because of its indirect benefits. Well-off urban dwellers, whose taxes finance the program, support it because it helps to dampen rural migration into Bombay. Rural elites benefit from the infrastructure that results (better roads, less erosion, reforestation) and from an assured supply of labor because the guarantee of work keeps the rural farm labor force in the area through the lean season. Finally, the guarantee of assistance in times

of need ensures the support of many community members who do not participate normal times. There would seem to be important lessons here for designing schem with built-in political sustainability.

The administrative costs in this case equal the scheme's nonwage costs (mater inputs and supervision). The size of these outlays is limited by the EGS chart which stipulates that wages must account for about two-thirds of the total cost the project.

Finally, costs may be associated with behavioral responses. Clearly participar would have been engaged in alternative activities had they not been employed the relief work program. Some of those activities would have been incom earning, and that income clearly must be netted out to calculate the net benefi But the counterfactual of what the participants would have done without the schen is difficult to estimate, and with the exception of one study by Ravallion and D (1995), no convincing attempts to do so exist for any public works program Their study relies on longitudinal data covering a ten-year period for two villag in which some 50 percent of the population was poor by local standards. To est mate forgone incomes, the authors devised a time allocation model to predict th time displaced from income-earning activities due to the EGS. They find relative low forgone incomes at about 25 percent of the wages earned from the relief wor program. Work on the program primarily displaced unemployment, leisure, an domestic activities.

How Did the Poor Fare?

Putting all the costs and the targeting performance together, the study calculates tha about 50 percent of the benefits go to the poor and that poverty is reduced. Yet, thi outcome is the same as that achievable if the same amount of public funding ha been distributed as an untargeted, uniform transfer to all households. The impact o poverty would be just as large. That result is attributable in part to the high inci dence of poverty and in part to the simplifying assumptions that the money can b uniformly distributed without administrative expense and without affecting behav ior. Still, this hypothetical benchmark shows that concentrating benefits on the poo does not necessarily maximize their gains because it may generate costs.

The key lesson is not that spending should be untargeted, but rather that the cas must rest on the net impact on poverty. Policymakers need to pay attention, fo example, to the costs associated with project design. In the case of the EGS, the poten tial indirect benefits must be assessed as well. Benefits from the assets created, re duced risks, and potential effects on other wages could be substantial. There is anec dotal evidence for all three, but little research to draw on. The case for the EGS ma ultimately rest on the size of its indirect benefits.

Conclusions

Targeting will often be desirable, but it should not be prejudged. Each case should be examined as carefully as feasible. The motivation for narrow targeting is often that broad targeting is costly, especially because of leakage. In trying to target narrowly, however, costs can be created that can wipe out the benefits. Both types of targeting have costs for the poor. Those incurred by narrow targeting are just more hidden and thus more often ignored. One can easily overestimate the gains to the poor from narrow targeting.

Both broad and narrow targeting may also have hidden benefits, notably when there are dynamic gains from relaxing the constraints that poor people face. Even less is known about benefits than about the costs. Knowledge about both costs and benefits suffers from inadequacies in data and methods of analysis and from the specific context of results. It is clear that the conventional static incidence picture is not going to capture the hidden costs and benefits from behavioral responses and impacts that occur over time. Conventional methods of analysis should thus not be the sole basis for decisionmaking. Reasonably reliable short-cut methods need to be pursued side by side with more rigorous and costly evaluations to test those methods.

The choice between broad or narrow targeting is not clear-cut. Ideally, targeted schemes should be designed so that incentives for escaping poverty are not destroyed and with better incentives both for participants and for their administrators. One should look for design features that encourage the poor to select themselves and do not discourage them from escaping poverty by their own means. The conditions conducive to the success of self targeting will not always be present, however. Furthermore, there may be no indicators that are fixed, easily observed, and good proxies for low incomes. The choice between broad or narrow targeting will depend on country circumstances including constraints on policy instruments, administrative and institutional. It will also depend on policy objectives. Indeed, although the appropriate combination will differ across countries, the two approaches will often be complementary. The best approach may often be a combination of broad targeting of social sector and basic infrastructure spending with narrow targeting of transfers for neglected groups and objectives.

Achieving better outcomes for poor people from public spending may not be easy, cheap, or rapid, and this fact may have implications for other areas of policy. For example, macroeconomic operations may have to be timed and sequenced to allow for the frequent inability to compensate losers in the short term, particularly when administrative capabilities for compensation are weak. Governments may also have to engage in institutional strengthening and capacity building. Building safety nets is a long-term investment; one cannot simply wait for a crisis—the institutional capacity must already be in place if there is any hope of responding in time.

Designing effective programs for the poor also requires collecting appropriate d: for understanding poverty and monitoring economy-wide, as well as targeted, pol impacts. Policymakers need to experiment widely, collect information, monitor co and outcomes of specific schemes, conduct impact evaluations, form careful asse ments of whether schemes are working, and be ready to revise and change them necessary. Careful, rigorous evaluations will not be logistically or economically pr: tical for each new intervention, but periodic, strategically chosen efforts can provi valuable feedback and lessons for design and implementation. There are almost c tainly no easy solutions.

Note

Dominique van de Walle is senior economist in the Development Economics Research Group. [author would like to thank Harold Alderman, Jennie Litvack, and Martin Ravallion for their us comments.

1. There is a sense in which policies that are typically categorized in the literature as broad targ ing are conceptually similar to indicator targeting when the indicator is demand for a certain type public service.

2. Geographical targeting had been examined for India (Datt and Ravallion 1993), Indon (Ravallion 1993), and several Latin American countries (Baker and Grosh 1994). Targeting landless in Bangladesh is explored by Ravallion and Sen (1995). Family allowances in Hungary explored by van de Valle, Ravallion, and Gautam (1994).

References

The word "processed" describes informally reproduced works that may not be commonly ava through library systems.

Alderman, Harold, and Lindert, Kathy. 1998. This volume.

Anand, Sudhir, and Ravi Kanbur. 1990. "Public Policy and Basic Needs Provision: Interven and Achievement in Sri Lanka." In Jean Drèze and Amartya Sen, eds., *The Political Econom Hunger*. vol. 3, *Endemic Hunger*. Oxford: Clarendon Press.

Baker, Judy L., and Margaret Grosh. 1994. "Poverty Reduction through Geographical Targe: Does It Work?" *World Development* 22(7):983-94.

BIDS (Bangladesh Institute of Development Studies). 1997. "Enhancing Accessibility to and Re tion in Primary Education for the Rural Poor in Bangladesh: An Evaluation of the Food Education Program." Dhaka. Processed.

Besley, Timothy. 1997. "Political Economy of Alleviating Poverty: Theory and Institutions Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Developmen nomics 1996*. Washington, D.C.: World Bank.

Besley, Timothy, and Stephen Coate. 1992. "Workfare vs. Welfare: Incentive Arguments for W Requirements in Poverty-Alleviation Programs." *American Economic Review* 82(1):249-61.

Besley, Timothy, Stephen Coate, and Timothy W. Guinnane. 1996. "Understanding the W house Test: Information and Poor Relief in Nineteenth-Century England." Yale Univers Department of Economics, New Haven, Conn. Processed.

Besley, Timothy, and Ravi Kanbur. 1993. "Principles of Targeting." In Michael Lipton and Jacques van der Gaag, eds., *Including the Poor.* Washington, D.C.: World Bank.

Cornia, Giovanni Andrea, and Frances Stewart. 1995. "Two Errors of Targeting." In van de Walle and Nead (1995).

Datt, Gaurav, and Martin Ravallion. 1993. "Regional Disparities, Targeting, and Poverty in India." In Michael Lipton and Jacques van der Gaag, eds., *Including the Poor.* Washington, D.C.: World Bank.

Demery, Lionel, Julia Dayton, and Kalpana Mehra. 1997. "The Incidence of Social Spending in Côte D'Ivoire." World Bank, Poverty Reduction and Economic Management Network, Washington, D.C. Processed.

Gelbach, Jonah, and Lant Pritchett. 1996. "Does More for the Poor Mean Less for the Poor? The Politics of Tagging." Policy Research Working Paper 1523. World Bank, Development Economics Research Group, Washington, D.C. Processed.

Grosh, Margaret. 1995. "Toward Quantifying the Tradeoff: Administrative Costs and Incidence in Targeted Programs in Latin America." In van de Walle and Nead (1995).

Hammer, Jeffrey, Ijaz Nabi, and James Cercone. 1995. "Distributional Impact of Social Sector Expenditures in Malaysia." In van de Walle and Nead (1995).

Jacoby, Hanan. 1997. "Self-Selection and the Redistributive Impact of In-Kind Transfers." *The Journal of Human Resources* 32(2):233–49.

Jalan, Jyotsna, and Martin Ravallion. 1998. "Are There Dynamic Gains from a Poor-Area Development Program?" *Journal of Public Economics* 67(1):65–85.

Jensen, Robert. 1998. "Public Transfers, Private Transfers, and the 'Crowding Out' Hypothesis: Evidence from South Africa." Princeton University, Department of Economics, Princeton, N.J. Processed.

Radhakrishna, R., S. Indrakant, C. Ravi, and K. Subbarao. 1996. "India's Public Distribution System: A National and International Perspective." World Bank, Poverty and Social Policy Department, Washington, D.C. Processed.

Ravallion, Martin. 1993. "Poverty Alleviation through Regional Targeting: A Case Study for Indonesia." In Karla Hoff, Avishay Braverman, and Joseph Stiglitz, eds., *The Economics of Rural Organization: Theory, Practice and Policy.* New York: Oxford University Press for the World Bank.

Ravallion, Martin, and Gaurav Datt. 1995. "Is Targeting through a Work Requirement Efficient? Some Evidence from Rural India." In van de Walle and Nead (1995).

Ravallion, Martin, and Binayak Sen. 1995. "Impacts of Land-Based Targeting on Rural Poverty." *World Development* 22(6):823–38.

Sahn, David E., and Harold Alderman. 1995. "Incentive Effects on Labor Supply of Sri Lanka's Rice Subsidy." In van de Walle and Nead (1995).

Subbarao, K., Aniruddha Bonnerjee, Jeanine Braithwaite, Soniya Carvalho, Kene Ezemenari, Carol Graham, and Alan Thompson. 1997. *Safety Net Programs and Poverty Reduction. Lessons from Cross-Country Experience.* Directions in Development Series. Washington, D.C.: World Bank.

Tuck, Laura, and Kathy Lindert. 1996. *From Universal Food Subsidies to a Self-Targeted Program: A Case Study in Tunisian Reform.* Discussion Paper 351. Washington, D.C.: World Bank.

van de Walle, Dominique. 1992. "The Distribution of the Benefits from Social Services in Indonesia 1978–87." Policy Research Working Paper 871. World Bank, Development Economics Research Group, Washington, D.C. Processed.

———. 1994. "The Distribution of Subsidies through Public Health Services in Indonesia 1978–87." *The World Bank Economic Review* 8(2):279–309.

———. 1998. "Assessing the Welfare Impacts of Public Spending." *World Development* 26(3):365–79.

van de Walle, Dominique, and Kimberly Nead. 1995. *Public Spending and the Poor: Theory an Evidence*. Baltimore: The Johns Hopkins University Press.

van de Walle, Dominique, Martin Ravallion, and Madhur Gautam. 1994. *How Well Does the Soci Safety Net Work? The Incidence of Cash Benefits in Hungary 1987–89.* Living Standards Measure ment Study Working Paper 102. Washington, D.C.: World Bank.

Wodon, Quentin. 1998. "Cost-Benefit Analysis of Food for Education in Bangladesh." World Bank Development Economics Research Group, Washington, D.C. Processed.

World Bank. 1997. *Bangladesh: Making the Best Use of Public Resources.* Washington, D.C.

Using International Institutions to Improve Public Procurement

Bernard Hoekman

The World Trade Organization's voluntary rules on government procurement are a useful mechanism for ensuring that public procurement procedures are efficient. They also provide an opportunity to reduce the uncertainty of participants by increasing transparency and accountability. Yet most developing countries have chosen not to subject their procurement policies to international disciplines and multilateral surveillance. Their reasons may include an unfamiliarity with the government procurement agreement (GPA); a perception that the potential payoffs are small; a desire to discriminate in favor of domestic firms; or the successful opposition of groups that benefit from the current regimes. Although the economic rationales for abstaining from the GPA are not compelling, a quid pro quo for accession may be needed to overcome opposition by special interests. Developing country procurement markets are large enough that governments may be able to make accession to the GPA conditional on temporary exceptions to multilateral disciplines or on better access to export markets.

All over the world government agencies purchase the goods and services necessary to provide the public with education, defense, utilities, infrastructure, public health, and so forth. The public procurement associated with these expenditures often represents a significant share of a country's gross domestic product. To maximize the use of scarce financial resources, governments have developed procedures and mechanisms to ensure that public entities procure these collective goods and services at least cost and in a fair and transparent manner. Some countries have addressed the issue of procurement efficiency head-on; South Africa, for example, wrote a section on procurement principles into its 1994 constitution requiring the government to pursue a fair, public, and competitive procurement process under the direction of independent and impartial tender boards that are obligated to record decisions and make them available to interested parties (Transparency International 1997).

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 249–69

Many procurement systems attempt to mimic the workings of the market by requiring that public entities seek competitive bids from potential suppliers of goods and services. Over time, an increasing number of governments have also pursued more far-reaching efforts to subject production units directly to competitive forces by privatizing state-owned enterprises, encouraging competitive entry into sectors traditionally reserved for the state (for instance, utilities), and by contracting out activities to the private sector. In a survey of the empirical literature on the impact of competitive tendering and outsourcing, Domberger, Hall, and Lee (1995) conclude that savings on the order of 20 percent are common and do not come at the expense of quality.

In the late 1970s several high-income countries negotiated an agreement on government procurement under the auspices of the General Agreement on Tariffs and Trade (GATT). The Government Procurement Agreement (GPA) extends the basic principles of the GATT—nondiscrimination, national treatment, and transparency—to the tendering procedures of specified government entities. Members of the GPA are Canada, the European Union, Hong Kong, Israel, Japan, Republic of Korea, Norway, Singapore, Switzerland, and the United States. Although many developing-country governments have made efforts to reduce the cost of providing public services, they have refrained from signing the GPA. This is surprising, because the GPA appears to be a useful mechanism for ensuring that government procurement procedures maximize value for money. The benefits are likely to include not only a reduction in procurement costs but, perhaps more important, positive spillover effects that result from transparency and accountability.

This article examines multilateral government procurement regulations from a developing country perspective. It assumes that governments are interested in achieving and maintaining efficient, transparent, and accountable procurement procedures and asks why developing countries have not signed on to the GPA. This is an important topic because the multilateral rules of the game must be consistent with "best practices." If there are good reasons to question the economics of the GPA from a development and growth perspective, the rules should be revised. If not, the issue becomes one of political economy. Efforts must then be directed at identifying and overcoming the resistance to membership by those groups in society that benefit from the status quo at the expense of the community at large.

Developing countries can expect to face increasing pressure on the subject of government procurement in bilateral trade relations with industrial nations. The United States has played a leading role in this connection, making public procurement practices a priority and linking them to the broader issue of combating corruption. In April 1996, largely at the initiative of the United States, members of the Organisation for Economic Co-operation and Development (OECD) agreed that firms would not be permitted to write off bribes against tax obligations; in May 1997 they recommended classifying bribery of foreign officials as a criminal offense under national

legislation (World Bank 1997). At the December 1996 meeting of the World Trade Organization (WTO) in Singapore, a working group was created to conduct a study on transparency in government procurement practices and develop "elements for inclusion in an appropriate agreement" (WTO 1996, p.7). Although the GPA is not mentioned, the most straightforward way for developing countries to deal with procurement-related concerns is to join the GPA. As far as the United States is concerned, this is clearly the ultimate objective; in the meantime, an interim agreement on transparency, openness, and due process is seen as "an important step toward a more comprehensive multilateral agreement in the WTO" (USTR 1997, p. 4).

The WTO Agreement on Government Procurement

The GPA applies to laws, regulations, procedures, and practices pertaining to any government procurement, including purchases, rentals, leases, and lease-purchase agreements, with or without the option to buy. The procurement agreement covers only entities listed by each signatory in three categories: central government; subnational entities; and others (these are primarily utilities that may be partially or wholly privately owned). Procurement is subject to the GPA as long as its value exceeds certain specified thresholds and the goods or services in question are not exempt from the Agreement. The threshold for central governments is generally Special Drawing Rights (SDR) 130,000; for local governments SDR 200,000; and for other entities usually SDR 400,000. For most members, construction contracts are covered only if they exceed SDR 5 million. (In 1997 one SDR was equivalent to $1.35.)

All procurement of goods is covered unless otherwise specified. (In most instances military procurement is the exception.) In the case of services, only those products that are explicitly listed by each member country are subject to the GPA's rules. In practice these commitments closely parallel those of the World Trade Organization members under the General Agreement on Trade in Services (GATS), which liberalizes access to service markets generally (see Hoekman and Kostecki 1995 for an introduction to the GATS and references to the literature).

General Provisions of the GPA

The primary obligations imposed by the GPA are transparency and nondiscrimination. This policy extends not only to imports but also to goods and services provided by local subsidiaries of foreign firms. Preferential prices, offsets (measures, such as domestic content requirements and technology licensing, that encourage local development), and similar policies that discriminate in favor of domestic firms are in principle prohibited; competitive tendering procedures are encouraged. These include open tendering, where any supplier may respond to a published call for ten-

ders, or selective tendering, where bids are restricted to prequalified suppliers who have demonstrated that they meet technical competence norms. Limited tendering under which potential suppliers are directly solicited to bid by the procuring entity may be used in only three circumstances: situations in which no responses have been received to an open or selective call for tenders, cases of urgency, and orders for additional deliveries by an original supplier.

Calls for open tenders must be published in all cases, stating the mode of procurement, its nature and quantity, dates of delivery, economic and technical requirements, amounts and terms of payment, and so on. Individual suppliers may not be given information that could have the effect of precluding competition. Entities are obliged to award contracts to the tenderer who "has been determined to be fully capable of undertaking the contract" and who is either the lowest cost or most advantageous supplier, according to the evaluation criteria set forth in the notices. Because an evaluation of which tender comes closest to satisfying the criteria is open to a considerable degree of discretion, much depends on how the criteria in the notices or tender documentation are worded. It is a violation of the procurement agreement to determine that a tender is the most advantageous on the basis of criteria that are not specified.

The nature of procurement is such that unless rapid remedial action can be taken to intervene in the procurement process, firms are unlikely to contest perceived violations of the rules A unique feature of the GPA is that it requires members to establish bid protest or challenge procedures, under which bidders can correct breaches of the GPA in order to preserve commercial opportunities. Such measures may involve suspension of the procurement process, reopening of the tender procedure, or the award of compensation for loss or damages. This is a key dimension of the GPA because it gives firms an incentive to defend their interests. The domestic challenge mechanism is complemented by the WTO's multilateral dispute settlement process. To ensure transparency and facilitate the application of these procedures, procuring entities must provide information explaining why a supplier's application to qualify was rejected; why an existing qualification was terminated; and why a tender was not selected. They must also identify the winning bidder and clarify the characteristics and relative advantages of the tender selected.

Developing Country Provisions

Although in principle the GPA prohibits signatories from discriminating in favor of domestic firms, developing countries may negotiate mutually acceptable exclusions from the rules on national treatment for certain entities, products, or services (Article V). Such negotiations may also be initiated after signing the procurement agreement. That option, however, is limited to *certain* entities, products or services, and the scope to pursue such policies is therefore inherently limited by the relative neg

tiating power of the country seeking to apply them. Developing countries may also, at the time of accession, negotiate conditions for the use of offsets. Although offset requirements explicitly allow for de facto discrimination against foreign suppliers, the requirements may be used only to qualify for the procurement process and not as criteria for awarding contracts. Thus, if a firm offers local content that greatly exceeds the minimum requirements, that aspect of the offer may not be a factor in awarding a contract.

Economic Issues

The major substantive disciplines imposed by the GPA are nondiscrimination, competitive tendering, and transparency, complemented by the domestic and multilateral enforcement mechanisms. What is the economic rationale for these rules? Are there reasons to conclude that they may not be in the interest of developing countries? There are two potential sources of concern: first, discrimination may be needed to ensure least-cost procurement; and second, the net economic payoffs associated with the rules may be too small.

Is Nondiscrimination Always Optimal?

Intuitively the nondiscrimination rule appears to be unambiguously beneficial because it should intensify competition and thereby minimize the costs of procurement. As is often the case in economics, this proposition is not necessarily true. Discrimination may be necessary to minimize average procurement costs. Discriminating against foreign bidders may be welfare improving *if* domestic firms are at a competitive disadvantage in certain areas (that is, if they are higher-cost producers) and only a limited number of firms (foreign and domestic) bid for a contract. In such situations, foreign firms may exploit their cost advantage by bidding just below the amount they expect domestic firms to bid (McAfee and McMillan 1989). Although the foreign firm will be the lowest bidder, the bid may be substantially above the firm's actual cost. A policy that gives preferences to domestic firms may then induce foreign firms to lower their bids by the extent of the preference margin. If so, procurement favoritism increases national welfare. Even if the cost structure of domestic and foreign firms are identical and account is taken of the social cost of distortionary taxation, discrimination may be rational simply because foreign profits do not contribute to domestic welfare (Branco 1994). Thus favoritism can be used as a rational profit-shifting strategy whenever there is imperfect competition. Shifting demand to domestic firms may also reduce price-cost margins as domestic output expands (Chen 1995).

Even when there are many potential suppliers for a contract, which should guarantee the lowest costs, discrimination may be beneficial in other ways. If, for in-

stance, the products to be procured are intangible or if enforcing contract com|
ance is a problem, public agencies may need to grant excess profits to contract
to get them to deliver (Laffont and Tirole 1991; Rothenberg 1993). Moreover
there is an incentive to pay a premium over the suppliers' cost to ensure contr
performance, the required premium may increase as the number of potential b
ders rises (because each supplier will take into account the higher probability
not getting repeat business) (Breton and Salmon 1995). Minimizing the costs
procurement in such settings may require limiting the number of potential sup
ers. If so, governments can be expected to favor domestic over foreign supplie
The sourcing costs will not be affected, while political benefits may arise beca
domestic firms are part of their constituencies. Such situations are more likely
the case of service procurement, because of their intangible nature. Services
often the largest category of government purchases—increasingly so in countr
that have been pursuing outsourcing and contracting strategies. In the United Stat
for example, most federal nondefense procurement is for services (Francois, Nels
and Palmeter 1997).

Problems of asymmetric information may also reinforce the decision to cho
local suppliers in order to reduce monitoring costs. Such proximity incentives ma
it more difficult for foreign firms to bid successfully, even when they are not fa
with formal discrimination, and may encourage them to contest procurement m
kets by investing directly in developing countries. This incentive is, of course, r
procurement-specific but applies in all instances where buyers prefer to deal wi
"local" suppliers. The policy issues then are how to decide whether suppliers are lo
"enough" and what—if any—barriers exist against foreign direct investment

Although discriminatory procurement may lower procurement costs in some si
ations, studies suggest that the net welfare benefits are likely to be modest at be
increased prices will tend to offset any cost savings (Deltas and Evenett 1997) I
net welfare impact depends on the government's objective function, particularly tl
relative importance given to domestic industry profits as opposed to expected p
curement costs. The latter will generally be a multiple of the former. In many situ
tions the information required to determine whether discrimination is beneficial
not likely to be available. Even if it is, the wording of general regulations calling f
discrimination—for instance, a universal price preference of 15 percent—is too rig
In many cases markets will be competitive and products will be relatively hom
genous, so these considerations do not arise.

Nonetheless, in principle there is a potential tradeoff between the GPA's nondi
crimination rules and economic efficiency. One way to address this tension would b
to give governments the discretion to apply a "rule of reason" and require them
rationalize their decisions. This approach, however, introduces the potential for ar
bitrary decisions and makes it difficult—if not impossible—to apply the GPA's en
forcement and dispute settlement provisions.

Experience with competitive procurement regimes indicates that in most situations competition is the best rule of thumb; competitive and transparent procurement regimes generate substantial cost savings. For example, Transparency International (1997) notes that noncompetitive procedures may increase procurement costs as much as 30 percent. Estimates of the cost savings associated with international competitive bids for World Bank loans are in the same range. In the case of a recent balance of payments loan to the Russian Federation financed by the World Bank, procurement costs under competitive bidding procedures were on average 30 to 40 percent below the costs of identical items acquired in the past without such bidding and as much as 75 percent less for certain pharmaceutical products (World Bank 1994).

In practice discrimination may be motivated not by considerations of cost or contract compliance but by a desire to promote domestic industry. Much has been written about the pros and cons of protecting infant industries. What matters in the current context is not whether there are rationales for intervention on infant industry grounds, but that procurement favoritism is unlikely to be the optimal policy to use in pursuit of this objective. In effect, it is equivalent to a subsidy financed by domestic taxpayers, because government purchasing costs increase and must be financed. But it is less transparent than a direct subsidy from the budget and as a result provides greater potential for rent-seeking and corruption. Debroy and Pursell (1997) provide an interesting review of India's use of procurement regimes to support infant industry policies—a policy the government recently has begun to reverse.

Costs and Benefits

Even if it is accepted that nondiscrimination is the best rule of thumb and that it will produce significant cost savings, the net benefits may be minor because of the costs incurred. These costs can be substantial, as entities must satisfy many procedural requirements to ensure due process and transparency. These procedures are undoubtedly burdensome, but they have advantages as well. Nondiscrimination, transparency, and accountability provisions may constrain rent-seeking activities. Allowing procuring entities to discriminate may facilitate bribery of procurement officials. Although the issue of corruption extends beyond procurement, rent-seeking in the public purchasing context is particularly prominent because the amounts involved are significant and foreign interests are frequently affected. Case studies have demonstrated that corruption can increase the costs of a project by as much as 25 to 50 percent (Wade 1982; Rose-Ackerman 1995a). Corruption and rent-seeking reduce economic growth, distort resource allocation, and result in higher taxes or, more commonly, deficit financing (Mauro 1995; Murphy, Schleifer, and Vishny 1993; Schleifer and Vishny 1993; Bardhan 1997).

Abstracting from differences in cultural norms across countries, effective anticor-ruption strategies must reduce the magnitude of the benefits that can be granted b officials, increase the costs of bribery for the private sector, and limit the mark power of officials (Rose-Ackerman 1995a, 1995b; Bardhan 1997). Of the variou strategies and suggestions offered in the literature, the following are particularly rel-evant for procurement: effective deterrents through ex post punishments that excee the gains realized (including banning firms caught in attempts to engage in briber from bidding for contracts for a number of years); the creation of external monitor-ing devices and institutions (including encouragement and protection of "whistle blowers"); public transparency mechanisms (published audits by independent audi-tors, a free press); privatization and hard budget constraints; requiring the use o standardized products and goods that have well-established market positions; the us of general retail-wholesale market prices for goods similar to those to be procured comparators; and the use of incentives to encourage bidders to complain if the suspect corruption.

Although the GPA lacks an explicit corruption standard or norm, it is consisten with—or embodies many of—these principles. Of particular importance are the chal-lenge procedures that allow firms to protest before the decision process is completed as well as thereafter. Multilateral monitoring and the threat of WTO dispute settle-ment procedures will also help to ensure that entities abide by the GPA's substantive and procedural disciplines. Alam (1995) argues that the opportunities for losers to take countervailing actions can constrain rent-seeking activities. In the context of procurement, the set of losers is usually small. In discretionary, nontransparent pro-curement systems, firms that lose out have little incentive to protest irregularities because they fear being blacklisted. The GPA's rules are designed to maximize the incentives to obtain and use information concerning possible violations, but give the sunk costs of participating in the bidding process, such a protest is only viable the expected returns outweigh the costs of protesting. Ensuring that this is the case may be difficult.

Operation of the GPA

The GPA requires signatories to report annual procurement statistics to the W Committee on Government Procurement, which provides oversight for the GPA This committee, which comprises representatives from all member countries, meet periodically in Geneva. Signatories began reporting statistics for the year 198 Because the latest year available for some countries is 1992, the analysis here is re-stricted to that period—a time during which the GPA applied only to the procure-ment of goods by central government entities. As a result, most procurement activi-ties were excluded. In the case of the United States, for example, Francois, Nelson

and Palmeter (1997) note that in 1993 goods accounted for less than 5 percent of total federal nondefense-related purchases.

Coverage

In 1992 the total procurement of goods by the entities covered by the GPA was about $62 billion. Average annual purchases that year by the United States, the largest procurement market, were some $29 billion (table 1). This compares to total purchases of $16 billion by the European Union (EU)-12 entities, $9.2 billion by Japan, and $1.6 billion by Canada. For the purposes of cross-country comparisons, it is helpful to relate these numbers to total central government expenditures on nondefense-related goods and services and to capital expenditures by the central government. Relative to total central government expenditures on goods and services—which includes items such as wages—purchases under the GPA by large countries such as France, Germany, Italy, Japan, and the United Kingdom tend to be below average; the United States and the Nordic countries are above average. On both measures, Germany and Italy have the lowest ratios, suggesting they may have sought to limit the coverage of their GPA obligations. Israel and Singapore have scheduled substantially fewer procuring entities than the OECD countries.

For a number of reasons it is difficult to estimate how much additional procurement was brought under the GPA's umbrella in 1996, when it was extended to cover services and subnational entities. First, it is not always clear whether specific services are covered or whether the nondiscrimination rule applies to all services. (Many countries made their commitments conditional upon reciprocity.) Second, no information on the size of the average contract and the types of goods and services purchased is available. The best estimates of central government nondefense expenditures by GPA signatories puts the total at about $2.1 trillion (table 1); subnational government bodies add at least another $1 trillion (IMF 1996). Not all of this procurement is available for international competition. In particular, the average subnational government contract is likely to be smaller than those of central government entities. If it is assumed that one-third of total outlays by central and subnational governments could be subjected to GPA rules, the total potential market would be close to $1 trillion a year. Applying the historical GPA rate of 50 percent for the share of procurement of goods that falls below the threshold value, some $500 billion could be open to international competitive bidding (table 2). This is likely to be an overestimate, however, as threshold values for subnational government entities and services contracts are significantly higher than those that apply to central government entities. Because the thresholds for any construction contracts are at least SDR 5 million, and as much as SDR 15 million in Korea and Japan, $300 billion is a more realistic figure.

Table 1. *GDP and Central Government Expenditures, 1992*
(millions of dollars)

Country	Gross domestic product	Total government expenditure (A)	Total nondefense expenditure (B)	Capital expenditures (C)	Value of total contracts under GPA	Value of total contracts relative to (A)	(B)	(C)
Austria	185,235	40,282	29,842	4,750	433	1.08	1.45	9.12
Belgium[a]	218,836	64,197	44,540	4,973	407	0.63	0.91	8.19
Canada	493,602	128,155	96,959	1,965	2,399	1.87	2.47	122.12
Denmark	123,546	47,219	35,782	1,539	1,646	3.49	4.60	106.92
Finland	93,869	30,020	23,045	1,912	834	2.78	3.62	43.58
France[b]	1,319,883	256,711	178,104	19,273	3,279	1.28	1.84	17.02
Germany[b]	1,789,261	215,669	149,629	35,832	2,055	0.95	1.37	5.74
Ireland	43,294	16,986	13,236	1,269	208	1.23	1.58	16.43
Israel	69,762	25,270	13,001	3,500	68	0.27	0.53	1.95
Italy[b]	1,222,962	335,439	232,725	17,983	1,994	0.59	0.86	11.09
Japan[b]	3,670,979	479,158	332,436	58,360	9,507	1.98	2.86	16.29
Luxembourg	12,638	2,394	1,784	364	34	1.43	1.92	9.44
Netherlands	320,290	92,698	75,644	6,505	1,281	1.38	1.69	19.70
Norway	112,906	25,399	19,276	1,255	775	3.05	4.02	61.76
Singapore	46,025	8,889	3,424	2,272	30	0.34	0.89	1.34
Sweden	220,834	75,172	61,158	3,386	1,162	1.55	1.90	34.31
Switzerland (1984)	90,649	8,973	5,784	1,096	252	2.81	4.36	22.99
United Kingdom	903,126	338,259	211,108	33,755	5,740	1.70	2.72	17.00
United States	5,920,199	1,037,354	631,924	56,354	29,120	2.81	4.61	51.67
Total		3,228,246	2,159,399	256,343	61,227			
Unweighted average						1.64	2.60	30.35

a. Average 1991 and 1992.

b. ... excluded because of data inconsistencies.

[illegible] 1983–92

Country	Value of covered procurement 1983–85	Value of covered procurement 1990–92	Share of limited tendering 1983–85	Share of limited tendering 1990–92	Share of procurement covered domestically 1983–85	Share of procurement covered domestically 1990–92	Share of procurement that falls above SDR threshold 1983–85	Share of procurement that falls above SDR threshold 1990–92
Austria	179.4	403.9	43.2	43.2	51.3	1.5	46.9	43.5
Belgium[a]	129.9	407.2	10.3	8.6	100.0	100.0	40.0	57.5
Canada	968.5	2,163.9	9.7	7.4	92.3	77.4	42.9	46.0
Denmark[a]	447.9	1,651.7	4.0	1.4	95.6	79.8	5.8	12.6
Finland	256.3	800.8	0.3	0.1	95.7	69.1	39.4	18.1
France[a]	937.9	3,089.5	33.3	29.6	97.6	97.2	35.1	85.5
Germany[a]	845.6	2,029.3	13.3	24.0	97.2	99.1	24.9	61.7
Hong Kong	154.5	348.6	22.1	28.2	5.8	3.2	67.7	71.2
Ireland[a]	47.8	185.8	1.8	5.6	100.0	84.8	20.6	29.1
Israel	30.1	67.9	5.0	3.3	13.7	25.6	84.0	91.1
Italy[a]	382.4	1,937.5	1.0	5.9	99.2	97.8	5.0	60.5
Japan	3,379.4	9,274.1	12.3	21.0	86.1	85.5	37.4	43.7
Luxembourg[a]	14.9	35.8	39.8	23.9	100.0	97.8	30.9	28.4
Netherlands[a]	547.1	1,407.7	16.8	13.5	98.7	86.4	12.1	37.7
Norway	384.7	721.4	7.1	8.3	48.6	40.4	44.5	58.9
Singapore	48.5	31.0	0.0	0.0	45.1	67.3	57.9	51.1
Sweden	624.4	1,200.5	4.0	4.8	43.7	43.6	27.9	36.9
Switzerland	265.0	806.2	25.1	27.7	35.7	38.3	43.8	39.8
United Kingdom[a]	1,329.7	5,375.0	13.7	6.6	99.4	98.2	29.3	49.0
United States	24,080.8	28,891.2	11.5	9.6	86.3	90.4	80.1	66.8
Memo: EU countries	7,967.2	16,119.4	19.0	15.0	98.3	94.9	21.2	54.1
Average unweighted			13.1	13.0	74.6	71.2	38.8	49.5
Average weighted by value of procurement			13.3	14.1	86.1	87.3	59.7	57.4

a. Average of 1984–85.

Source: WTO (1984–94).

Foreign Sourcing

Smaller countries tend to procure more goods and services on international markets than do large countries. If Canada, the European Union, Japan, and the United States are excluded, about 60 percent of the purchases of goods by central government entities under the GPA are from national suppliers, compared with more than 90 percent for the large players. Unfortunately, EU statistics define "domestic" as within the European Union, thus skewing self-sufficiency ratios (see table 2). In interpreting these statistics, note that no distinction is made between domestic firms and foreign firms that have established a local presence. As long as the share of foreign direct investment is the same in large countries as in smaller ones, cross-country comparisons should not be affected.

In Japan, the United States, and the EU countries, the share of total procurement from domestic firms was virtually unchanged during 1983–92. The average weighted share across all GPA members is relatively constant; the unweighted average fell by 3 percentage points, from 74.6 percent to 71.2 percent (see table 2). Among smaller countries, however (with the exception of Singapore and Switzerland), the share of procurement from domestic sources has declined over time. Although such changes cannot be attributed solely to the GPA because other developments—such as the North American Free Trade Agreement, the EU procurement liberalization, and unilateral deregulation and privatization policies also played a role, the finding appears to be robust. But for most of the large players there has been no change.

This conclusion is supported by the few empirical studies on government procurement practices in OECD countries. These use a methodology suggested by Baldwin (1970) and Baldwin and Richardson (1972), which assumes that in the absence of discriminatory policy, government entities would behave in the same way as do private firms. Thus, government imports of a good as a share of total consumption would equal that of the private sector as a whole. The difference between the private sector import propensity and the actual import share of total government consumption can be called a preference margin. (A positive preference margin implies that the government share is lower than the private share.) Baldwin (1970) estimated that the preference margin in the United States was some 20 percent in 1958. After adjusting for the fact that certain large import items such as oil were not subject to discriminatory policies, the margin for the residual set of covered goods increased to some [illegible] percent. More recent estimates (1992) show the U.S. preference margin was 16 percent (Francois, Nelson, and Palmeter 1997). On a sectoral level, positive margins in OECD countries may be as high as 50 percent (table 3). Margins are invariably the highest for procurement of services.

The Baldwin-Richardson methodology is obviously sensitive to the assumption that, other things being equal, the government would import the same share of

Table 3. *Estimated Preference Margins for Core Government Purchasing, 1992*
(Baldwin-Richardson 1972 Approach)

Country or region	*Machinery*	*Other goods*	*Trade, transport, communication*	*Utilities*	*Other services*
Canada	—	—	—	—	39.6
United States	18.4	17.9	—	18.8	42.6
Western Europe	—	9.2	13.7	14.9	48.3
Japan	—	32.0	26.2	34.0	46.6
Australia	49.8	49.7	—	—	41.5
New Zealand	13.9	19.7	49.8	—	50.0
Korea, Republic of	30.6	20.8	—	—	48.2

— Less than or equal to zero.
Source: Francois, Palmeter, and Nelson (1997)

good as the private sector, and that all differences can be attributed to formal or informal preference policies. One source of bias that arises in this connection is that private sector imports may be distorted because of tariffs and other trade policies. Alternatively, private sector demand for certain products may be very low or even zero because government is the dominant supplier of output that embodies particular products (defense, utilities, certain types of transport services, and so on).

Limited Tendering

Limited tendering procedures (which involve an entity contacting and negotiating with potential suppliers individually) are permissible only under certain conditions. The use of limited tendering varies across signatories, from a reported low of zero (Singapore) to a high of more than 30 percent on average for France, Hong Kong, Italy, and Switzerland (see table 2). Across all signatories the average share of limited tendering was about 13 percent. This practice has been falling over time in the EU-12 (by about 7 percent a year on average during 1983–92), but rising in the United States. By 1992 limited tendering in the EU and in the United States stood at 10 percent, largely as a result of a significant decline in this practice by France and Germany during the 1980s. Japan's limited tendering rose from around 12 percent during 1983–85 to 21 percent during 1990–92. Hong Kong and Switzerland make even more intensive use of such procurement mechanisms.

This type of data is not available for developing countries. Indeed, little is known about the product composition of procurement across developing countries, the average contract size, or the "import propensity" of government entities. Indeed, one benefit of membership in the GPA would be that data on procurement flows would

have to be collected and reported to the WTO. This information would be valuabl
for researchers as well as for the governments concerned.

Transparency, Disputes, and the Challenge Mechanism

If procurement procedures are to function efficiently, potential suppliers need to b
aware of demand and have sufficient time to respond to calls for tenders. The GP
spells out these requirements, but it does not monitor compliance; that is left to th
private sector. As a result, violations may be overlooked. Even in the EU, where pro
curement opportunities must be published in the *Official Journal of the Europea
Communities* and can be obtained through electronic networks, a recent study foun
that procuring entities often failed to publish the information or provided insuff
cient time for responses. Moreover, many potential suppliers did not routinely monit
the *Official Journal* and were thus unaware of the potential market (Gordon, Rimmer
and Arrowsmith 1997). In part these problems result from ambiguities in the word
ing of the regulations on EU procurement.

The GPA's challenge mechanism is untested as yet, because it only came into forc
in 1996. The key requirement for these procedures to be effective is that participan
can either obtain very rapid intervention by a judicial or administrative body or hav
the opportunity to obtain significant financial compensation after the fact for viola
tions of procurement disciplines. (Once a tender has been closed and a contrac
awarded, reopening the proceedings may be difficult.) A concern with this process
its potential to give rise to problems of moral hazard and excessive litigation. Ensur
ing that firms have access to an effective and speedy mechanism through which the
can challenge a procurement process or award is therefore very important. A rec
evaluation of EU procedures and disciplines—which were an important model f
much of the GPA—suggests that EU suppliers have not used the challenge mechan
to great effect because it is perceived as being too slow (Gordon, Rimmer an
Arrowsmith 1997).

Negotiating Accession: Considerations of Political Econom

As noted earlier, membership in the GPA did not alter the sourcing patterns of larg
countries, but small countries purchased more from foreign suppliers over time. Thes
findings are consistent with economic theory. The larger the country, the greater th
number of potential domestic suppliers and the higher the probability that domest
firms attain minimum efficient scale. Thus, large countries can be expected to co
tinue to buy predominantly from national suppliers even if they abide fully by th
GPA. Small countries, in contrast, will generally have fewer national firms that ca
provide the goods needed by the government at least cost, especially speciali

capital-intensive items where scale economies are important (such as telecommunications, transport, or power-generating equipment).

The economic rationale for refusing to join the GPA is weak and may in large part be driven by political economy factors. The issue then is to seek to offset the opposition of groups that would like to maintain the status quo. The problem is similar to that facing trade reformers: incumbent industries and procurement officials have more of an individual incentive to oppose reform than losers from the status quo (taxpayers, consumers) have to push for it. In the context of trade liberalization, the problem arising from the asymmetric distribution of costs and benefits can be addressed by providing those who stand to gain from reform a greater incentive to engage in the political process in pursuit of their interests. One way to mobilize support for liberalization is to promise exporters better access to foreign markets, something that the government can demand as a quid pro quo from its trading partners for liberalizing the domestic market. But such access will not have much of an impact in increasing developing countries' share of OECD procurement markets. Although some countries may have a comparative advantage in providing certain services, service industries are much more subject to restrictive regulations than goods markets. There is no formal constraint on demanding other concessions, however. Developing countries could bargain for access to markets that are of interest to them. Although trade barriers are relatively low in industrial countries, a significant amount of tariff escalation remains. Tariffs and other trade barriers tend to be substantially higher on labor-intensive products such as footwear or clothing than on other types of manufactures. Achieving reductions in these barriers is difficult. Linking membership in the GPA to reductions in trade barriers by OECD countries may help mobilize support for domestic liberalization of procurement markets.

Another way to attenuate opposition is to change the rules of the game in a manner that meets some of the concerns associated with the implementation of the GPA. For example, developing countries could be allowed to apply price preferences that favor procurement from national suppliers. Although a strong presumption exists that the GPA's nondiscrimination, transparency, and enforcement provisions will benefit developing countries, many nations have procurement regimes that provide preferences for domestic firms. Such provisions may be motivated by infant industry arguments, they may reflect an attempt to offset the higher input costs of domestic bidders that result from other policies (for example, high import tariffs); or, as noted earlier, they may be a response to imperfect competition or asymmetric information. In principle, the optimal policy to deal with distortions such as tariffs is to eliminate or significantly reduce them, thereby eliminating domestic preferences as well. Some countries have embarked on a such a process, but local content requirements and price preferences continue to prevail in many jurisdictions.

The GPA could be amended to allow developing countries to establish price preferences as long as these are transparent and accountable. Although the case for policies

Table 4. *Government Expenditures and the Role of Official Development Assistance*
(millions of dollars)

Country group	*GDP 1992*	*Total nondefense expenditure on goods and services*	*Capital expenditure*	*Multilateral development bank loans, 1992*[a]	*Development Assistance Committee loans, 1992*[b]	*Ratio multilateral development bank loans/total nondefense expenditures*
Average of low-income countries (N=46)	23,704	2,053	1,637			
Total low-income countries	1,090,384	94,447	75,296	17,076	32,931	18.1
Average of lower middle-income countries (N=42)	31,151	2,237	921			
Total lower middle-income countries	1,308,362	93,960	38,682	13,972	15,337	14.9
Average of upper middle-income countries (N=16)	80,384	6,602	1,496			
Total of upper middle-income countries	1,286,158	105,631	23,943	6,135	1,239	5.8

a Refers to loan approvals

b Development Assistance Committee data pertains to net disbursements and includes bilateral assistance

Source: World Bank, Asian Development Bank, Inter American Development Bank, African Development Bank and OECD Development Assistance Committee, Annual Reports (various years), IMF (1996).

that discriminate in favor of domestic suppliers is not compelling, such an approac may be a useful transition to encourage membership in the GPA. Multilateral finan cial institutions, including the World Bank allow such preferences subject to certai conditions and limits. The World Bank permits a maximum price preference of 1 percent for procurement of goods, and 7.5 percent for public works projects. Provi sions for the use of price preferences are also included in the Model Law on Procure ment of the UN Committee on International Trade Law (Beveglia-Zampetti 1997) An advantage of price preferences over other policies that favor domestic procure ment is that they permit the price mechanism to continue to work. As such, they ar superior to local content and offset requirements, policies that the GPA currentl allows developing countries to apply. One option would be an agreement to conver offset and similar policies into price preferences. This would be akin to the conver sion of quotas and related measures to tariffs in the Uruguay Round for agricultur (see Hoekman and Kostecki 1995). Even if this meant that price preferences in som countries were relatively high, they would become the focal point for future multi lateral negotiations to reduce discrimination. Of course, the use of such preference should be optional and should not be used to create a preference margin that exceed the level currently implied by existing policies. Nor should tariff conversion be con sidered in instances where preference policies do not already exist.

Potential Negotiating Leverage

What negotiating leverage could developing countries exercise to pursue better ac cess to foreign markets or temporary exceptions to GPA disciplines? Given the mer cantilist nature of bargaining in the WTO context, the size of nonmember (develop ing country) procurement markets is perhaps the best indicator of leverage. Unfortu nately, the data on the size of procurement markets are very poor. Total centra government expenditures on goods and services by non-OECD countries in the earl 1990s was some $300 billion (table 4). This figure includes conservative estimate for several large countries that do not report data, including India, Indonesia, Paki stan, and South Africa.[1] Data on expenditures by subnational government entities i developing countries are even patchier. If it is assumed that the purchases by suc entities are equal to central government expenditures, the total will be at least $60 billion. How much of this is available to foreign suppliers depends on the share o contracts that exceeds the GPA's threshold values. These thresholds are likely to hav a greater effect in developing countries as the average size of each contract can b expected to be less than in a high-income country.

It is also important to recognize that many expenditures by developing countr governments are financed through official development assistance funds, both bilat eral and multilateral. Official bilateral development aid is often tied to buying good and services from the donor country; and the recipient government cannot subjec

aid-financed purchases to international competition. Absent an agreement fron nors to eliminate tied aid, such projects would be exempt from GPA rules. (versely, procurement financed through multilateral development assistance us already is subject to international competitive bidding.

Available data on the relative importance of aid flows as a share of govern expenditure suggest that aid finances a significant share of the total purchas goods and services by developing country governments (equivalent to 35 perce the total in low-income countries, 16 percent in lower middle-income nations, percent in upper middle-income economies; see table 4). This suggests that mu the procurement by poor countries either cannot be subjected to the GPA or ready subject to international competitive bidding. In practice, however, the expected to interest the higher-income developing countries in which aid plays a minor role. It can also be noted that what matters in the WTO context is not the actual policy stance of a government, but also the extent to which countries (lock in) their policies by making formal commitments not to become more re tive in the future. Thus, even if countries are already relatively open, this doe mean they have nothing to offer. An agreement to lock in a policy regime tl already nondiscriminatory has value in the WTO setting. Negotiating leverag developing countries is therefore substantial.

Conclusion

Government procurement is a key interface between the public and private sec Not only is public purchasing of fundamental importance in ensuring that gov ment gets the best value for its money, but procurement practices figure pro nently in the way potential investors and civil society at large view a country. A developing countries have adopted procurement legislation and regulations that to ensure that public entities purchase goods and services through an open and petitive process. To what extent actual practice is consistent with the formal and principles is often difficult to determine, in part because the incentives to test violations of the formal rules of the game are often small. The GPA provi unique international mechanism through which governments can credibly com themselves to a transparent and competitive procurement regime and provide ticipants with more effective enforcement mechanisms than may be available un the status quo. Enforcement is a necessary condition for any set of rules to be ef tive. This applies as much to industrialized as to developing countries.'

Governments desiring to minimize procurement costs have options that go yond the adoption of efficient purchasing practices and related procurement regi These include the privatization of government operations and private sector part pation in markets presently serviced by public entities. The potential benefits of

initiatives are increasingly recognized. Procurement disciplines may be a second-best option—if not redundant—if market forces can be applied to the activities of public entities. The greater the extent to which public entities operate in a competitive environment and are subject to hard budget constraints, the less likely that inefficient purchasing decisions will be made. Where public entities have a dominant position, however, competition laws and policies are required to ensure that they do not abuse their market power. The same is true of collusion and other restrictive business practices, matters on which the GPA has nothing to say (Wood 1997). This does not imply that procurement regimes will become irrelevant; quite the contrary. Governments will remain important purchasers of goods and services. Indeed, the more activities are contracted out by government agencies, the greater the need for transparent and efficient procedures to allocate contracts.

More research is required to determine the extent to which current procurement regimes in developing countries are compatible with the GPA. Surprisingly little is known about actual practices and their economic impact. Examples include data on the composition of procurement, the import propensity of government entities compared with those of the private sector, the extent to which procurement costs diverge from market price-based comparisons, whether differences in ownership and market structure across countries affect procurement efficiency of similar entities or contracts, and how frequently negotiations are reopened *after* contracts have been awarded.[3]

For developing countries, the main quid pro quo for membership in the GPA is reciprocal access to the procurement markets of existing members. This is not a very compelling incentive. Seeking better access to export markets more generally appears to be the best approach, complemented by efforts to seek transitional periods in which to phase in those aspects of the GPA, such as challenge mechanisms, where institutional strengthening will be required.

Notes

Bernard Hoekman is senior economist with the Development Research Group of the World Bank and research fellow at the Centre for Economic Policy Research in London. He would like to thank Marie Hélène Le Manchec and Faten Hatab for excellent research assistance; Annet Blank and Dick Wester for help in obtaining data, and Simon Evenett, Garry Pursell, and Alan Winters for helpful comments.

1. Data reported in IMF (1996) on the average ratio of expenditure on goods and services to total central government expenditures of all the developing countries reporting both variables was used to estimate data for missing countries.

2. In the EU, for example, many members states have been slow to implement European Community directives relating to procurement, and the directives are often violated. In 1994 the Commission of the European Communities initiated some 250 legal actions against members states or entities for inconsistency with procurement rules (*Financial Times*, November 23, 1996).

3. This is potentially a significant loophole in any procurement regime because it may reflect or [illegible] corruption and collusion between the entity and a favored supplier, who will bid low to win

the tender in the knowledge that the contract will be renegotiated ex post. This is alleged to prevalent practice in some European countries that belong to the GPA (*Financial Times*, Nover 23, 1994, p. 2).

References

The word "processed" describes informally reproduced works that may not be commonly avai through library systems.

Alam, M. S. 1995. "A Theory of Limits on Corruption and Some Applications." *Kyklos* 48:419

Baldwin, Robert. 1970. *Nontariff Distortions of International Trade*. Washington, D.C.: Brook Institution.

Baldwin, Robert, and J. David Richardson. 1972. "Government Purchasing Policies, Other and the International Monetary Crisis." In H English and K Hay, eds. *Obstacles to Trade in Pacific Area*. Ottawa: Carleton School of International Affairs

Bardhan, Pranab. 1997. "Corruption and Development: A Review of Issues." *Journal of Econ Literature* 35:1320-46.

Beveglia-Zampetti, Americo. 1997. "The UNCITRAL Model Law on Procurement of Goods (struction and Services." In Hoekman and Mavroidis, eds.

Branco, Fernando. 1994. "Favoring Domestic Firms in Procurement Contracts." *Journal of Inte tional Economics* 37:65–80.

Breton, Albert, and Pierre Salmon. 1995. "Are Discriminatory Procurement Policies Motivate Protectionism?" *Kyklos* 49:47–68.

Chen, Xiangqun. 1995. "Directing Government Procurement as an Incentive of Production *nal of Economic Integration* 10:130-40

Debroy, Bibek, and Garry Pursell. 1997 "Government Procurement Policies in India " In Hoek and Mavroidis, eds.

Deltas, George, and Simon Evenett. 1997. "Quantitative Estimates of the Effects of Prefer Policies." In Hoekman and Mavroidis, eds

Domberger, Simon, Christine Hall, and Eric Ah Lik Lee 1995 "The Determinants of Pric Quality in Competitively Tendered Contracts " *Economic Journal* 105:1454 70

Francois, Joe, Douglas Nelson, and David Palmeter. 1997. "Government Procurement in the A Post-Uruguay Round Analysis." In Hoekman and Mavroidis, eds

Gordon, Harvey, Shane Rimmer, and Sue Arrowsmith. 1998 "The Economic Impact of Regime on Public Procurement: Lessons for the WTO " *The World Economy* 21(2, March 87

Hoekman, Bernard, and Michel Kostecki. 1995. *The Political Economy of the World Trading S* Oxford: Oxford University Press.

Hoekman, Bernard, and Petros C. Mavroidis, eds. 1997. *Law and Policy in Public Purchasing WTO Agreement on Government Procurement*. Ann Arbor: University of Michigan Press

IMF (International Monetary Fund). 1996. *Government Finance Statistics Yearbook* Washin D.C.

Laffont, J. J., and Jean Tirole. 1991. "Auction Design and Favoritism." *International Journ Industrial Organization* 9:9–42.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 109:681 712

McAfee, R. Preston, and John McMillan. 1989. "Government Procurement and International Trade." *Journal of International Economics* 26:291–308.

Murphy, Kevin, Andrei Schleifer, and Robert Vishny. 1993. "Why is Rent-Seeking So Costly to Growth?" *American Economic Review* 83:409–14.

Rose-Ackerman, Susan. 1995a. "The Political Economy of Corruption." *Viewpoint* 74 (April).

———. 1995b. "Redesigning the State to Fight Corruption." *Viewpoint* 75 (April).

Rothenberg, Julio. 1993. "Comment." In Jim Leitzel and Jean Tirole, eds., *Incentives in Procurement Contracting*. Boulder: Westview Press.

Schleifer, Andrei, and Robert Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 107:599–617.

Transparency International. 1997. *The TI Source Book: Applying the Framework*. Internet address: http://www.transparency.de.

USTR (United States Trade Representative). 1997. "Annual Report on Discrimination in Foreign Government Procurement." April 30. Washington, D.C. www.ustr.gov/reports/special/title7.html.

Wade, Robert. 1982. "The System of Administrative and Political Corruption: Canal Irrigation in South India." *Journal of Development Studies* 18:287–327.

Wood, Diane. 1997. "The WTO Agreement on Government Procurement: An Antitrust Perspective." In Hoekman and Mavroidis, eds.

World Bank. 1994. "Rehabilitation Loan: Preidentified Import Component Distribution Monitoring Report." Russia Department, Washington, D.C. Processed.

———. 1997. *Helping Countries Combat Corruption: The Role of the World Bank*. Washington, D.C.

World Trade Organization. 1984–94. "Annual Reports to the Government Procurement Committee." Geneva.

———. 1996. "Singapore Ministerial Declaration." Geneva. Processed.

New Models for Old-Age Security: Experiments, Evidence, and Unanswered Questions

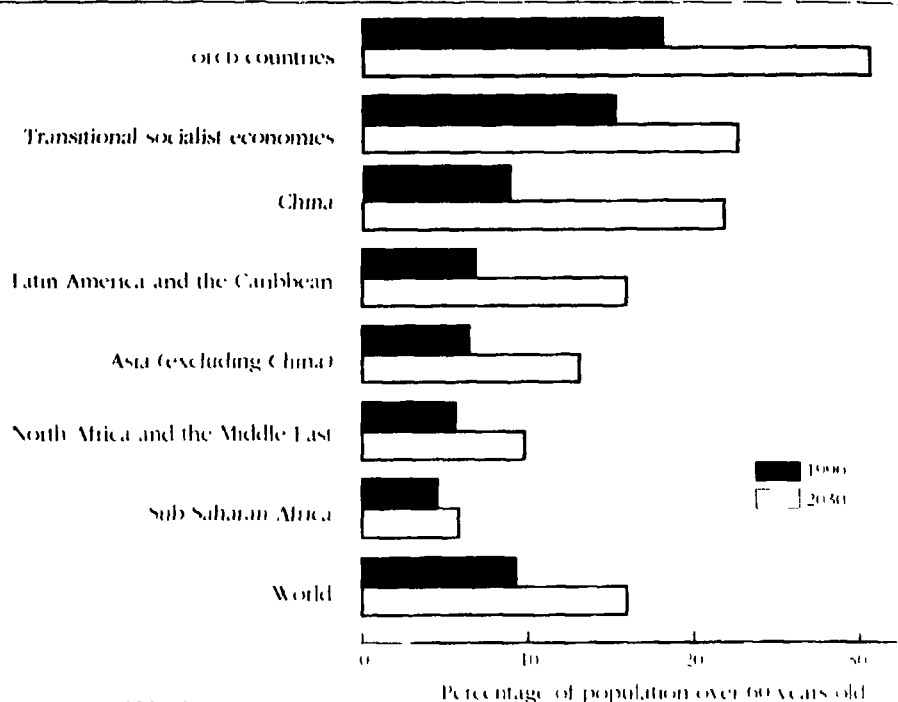
Estelle James

The escalating costs of traditional social security systems are forcing countries to reevaluate their formal programs that provide income maintenance support to the aging. This article suggests a reform strategy built around three systems, or "pillars," to provide old-age security—a public pillar with mandatory participation, a private, mandatory savings plan, and a voluntary savings system. Three variations of this model are being implemented in different countries: the Latin American model, in which individual workers choose an investment manager for their retirement funds, the OECD model, in which employers, union trustees, or both choose the investment manager for an entire company or occupation, and the Swedish notional account model, a reformed pay-as-you-go first pillar that may be supplemented by a second, funded pillar. Preliminary empirical evidence on the efficiency and growth effects of pension reform, mostly from Chile, indicates that the impact on national saving and financial market development and, through these, economic growth, has been positive and possibly large. Problems concerning high administrative costs and regulations that distort investment decisions remain to be resolved, however.

In the next 35 years the proportion of the world's population that is more than 60 years old will nearly double, from 9.5 to 16 percent (figure 1). With rapid increases in life expectancy and declines in fertility rates, the population in developing countries is aging much more rapidly than are populations in industrial countries. By 2030, as today's young working-age people near retirement, 80 percent of the elderly will live in what today are developing countries, and it is essential that policymakers start to plan for the care of their aging populations.

Many of these countries now rely heavily on family assistance and private, voluntary old-age support. But cross-sectional analysis shows that public spending on for-

The World Bank Research Observer, vol. 13, no. 2 (August 1998), pp. 271–301

Figure 1. *Percentage of the Population over 60 Years Old, by Region, 1990 and 20*

Source: World Bank (1994).

mal pension plans increases exponentially as populations age (figure 2). In so
industrial countries, for example, it now exceeds 15 percent of gross domestic pr
uct (GDP) and will soon reach that level in many more countries as the demograp
transition proceeds.

With such large sums involved, the way this money is generated and spen
affect the entire economy by influencing productivity, factor supplies, and there
the size of the GDP. For example, high payroll taxes can lead to unemploym
deficit financing can fuel inflation, and prefunding pension expenditures can be
of a plan to increase national saving. Countries with larger private pension p
have lower public spending, and these two types of spending may have diffe
effects on the broader economy (figure 3). Therefore, two criteria should be use
shape and evaluate these programs: they should protect the old in an equitable w
and they should promote, or at least not hinder, economic growth —which is impo
tant for both the old and the young.

In the past most government old-age security systems were pay-as-you-go pla
workers were taxed today to pay pensions to old people today. This article arg

that because aging is a predictable life experience, saving during younger years c self-insure a large part of old-age security, shifting consumption from younger pı ductive years to older years when consumption exceeds income. Myopia amo workers may require that retirement saving be mandatory, but relying to some (tent on self-insurance and saving for old age may reduce many of the incentive pro lems associated with tax and transfer pay-as-you-go systems.

Another aspect of old-age security systems requires pooling risks and insuring redistributing across individuals because some people will retire early with disabı ties, die young and leave dependents, live longer than average and run out of ı sources, or earn very low lifetime incomes that are insufficient to support them dı ing their working and nonworking lives. This is the rationale for providing combination of mandatory self-insurance and insurance across individuals in multipillar system that puts greater emphasis on saving, has separate financing a managerial mechanisms for redistribution and saving, and shares responsibility b tween the private and public sectors.

During the past few years, many countries have indeed been adopting multipill old-age security systems. Although structural change is always difficult, the experien of these countries shows that it is possible, that it takes somewhat different forms different places, and that it usually involves transition costs that are spread over seve generations. Preliminary empirical evidence suggests a positive impact on efficien and growth. But it also brings to the fore new problems --high administrative co and regulatory regimes that distort investment—that remain to be solved.

Problems with Old Systems

Most formal systems of old-age security today are publicly managed, pay "defin benefits" (meaning a payout formula based on the worker's earnings and year service), and are financed by payroll taxes on a pay-as-you-go basis. It is now wid recognized that these systems generate many problems, including

- High and rising payroll taxes that may increase unemployment
- Evasion and escape to the informal sector, where productivity is lower
- Early retirement, which reduces the supply of experienced labor
- Misallocation of public resources as scarce tax revenues are used for pensi rather than for education, health, or infrastructure
- Lost opportunities to increase long-term saving
- Failure to redistribute to low-income groups
- Unintended intergenerational transfers (often to high-income groups)
- The growth of a large implicit public pension debt and financing gap that make the current system unsustainable in many countries.

As a result, existing systems have not always protected the old and are particularly unlikely to protect those who grow old in the future. Moreover, they often have failed to distribute benefits in an equitable way and have hindered economic growth. In addition, they simply are not sustainable in their present form. Not each problem exists in every country, but they are found in most countries, both industrial and developing. This prevalence suggests that these problems are not accidental, but inherent in the economics and politics of pay-as-you-go defined benefit plans—the model preferred by politicians who find it simpler to promise short-term benefits at the expense of large long-term costs.

The Multipillar System

To avoid these dangers, the World Bank has been recommending and many countries have been moving toward a system in which some of an individual's pension is financed by preretirement savings, which are privately managed. Specifically, these new arrangements contain three pillars:

- A mandatory, publicly managed, tax-financed pillar for redistribution
- A mandatory, privately managed, fully funded pillar for savings
- A voluntary pillar for people who want more protection in their old age.

The first pillar resembles existing public pension plans, but it is smaller and focuses on redistribution—providing a social safety net for the old, particularly those whose lifetime income was low. The benefit formula can be flat (uniform for everyone or related to years of covered employment, as in Argentina and the United Kingdom), it can be means- and asset-tested (as in Australia), or it can provide a minimum pension guarantee (as in Chile). The first alternative provides additional co-insurance and redistribution to lower-middle-class workers while the last option is obviously cheaper. In some cases (Australia, Chile), the first pillar is financed out of general revenues rather than through a payroll tax. Because this pillar is of limited scope and has a broad tax base, the tax rates needed to support it are much lower than the public system requires in most countries today.

The second pillar differs dramatically from traditional systems. It links benefits actuarially to contributions as in a defined contribution plan, is fully funded, and is privately and competitively managed. (In such a plan the contribution is defined, and the future pension depends on accumulated contributions plus investment returns. In a fully funded system assets are always sufficient to cover future liabilities.) Essentially, people are required to save for their old age.

A third pillar, voluntary saving and annuities, offers supplemental retirement income for people who want more generous old-age pensions.

The most innovative and controversial of these arrangements is the second pill, so it is worth examining the rationale for its characteristics:

- Why mandatory? The rationale here is myopia—a significant number of peop may be shortsighted, may not save enough for their old age on a voluntary bas and may become a burden on society at large when they grow old.
- Why defined contribution? The close link between contributions and benefits this plan should discourage evasion, escape to the informal sector, and oth labor market distortions because people are less likely to regard their contri bution as a tax. And those who do evade bear the cost in the form of low benefits rather than passing the costs on to others and undermining the financi viability of the scheme. Because the pension is acquired on actuarially fair term given the age and accumulation of the worker, these plans are likely to deter ear retirement and to raise the normal retirement age automatically as longev increases—without involving the government in a difficult political decision
- Why fully funded? First, prefunding makes the costs clear up front so th countries are not tempted to make promises today that they will be unable keep tomorrow. Second, it avoids large payroll tax increases that are needed pay-as-you-go system as populations age. Third, it prevents large, inadverte intergenerational transfers from young people to older workers. Once unfunded system is set in motion, intergenerational transfers occur auto matically as a result of the aging and maturation process, sometimes in ways people did not expect and would not have chosen. For example, the ea generations to be covered (including its rich members) gain, while la generations (including its poor members) lose, even though they did not hav chance to participate in the political decision that produced this contract. Fu funding eliminates such undesirable transfers. And finally, funding may be used to help build long-term national savings. These savings can enhance productivity of future workers, they can be embedded in consumer durables t provide a stream of future services, and they can be invested abroad and redeemed to finance purchases of consumer goods. Thus, saving can be important ingredient of a long-term strategy for providing additional dome consumption when the dependency rate increases.
- Why privately managed? This maximizes the likelihood that economic rat than political objectives will determine the investment strategy, thereby pr ducing the best allocation of capital and the highest return on savings. And helps countries, especially middle-income countries, develop their financ markets. Empirical data show that publicly managed pension reserves typical earn low, even negative, returns, largely because public managers are required invest in government securities or loans to failing state enterprises at low nomina interest rates that become negative real rates during inflationary periods

hidden and exclusive access to these funds makes it easier for governments to run large deficits or to spend more wastefully than they could if they had to rely on a source of funds for which they were more accountable.

Competitively managed, funded pension plans, in contrast, are more likely to be invested in a mixture of public and corporate bonds, equities, and real estate, thereby earning a higher rate of return. Private pension funds can enjoy the benefits of investment diversification, including international diversification, which protects them from inflation and other country-specific risks and thus enables them to increase their yield and reduce their risk. Private pension funds build constituencies that help them resist political manipulation. They spur financial market development by creating a demand for new financial instruments and institutions. Three caveats obtain, however: countries must have at least rudimentary capital markets before they can put the funded pillar in place; considerable government regulation and regulatory capacity are needed to prevent fraud and excessive risk; and if this regulation is excessive or misdirected, financial markets and investment policies will not be optimal.

All three pillars co-insure against the many risks that old people face, particularly the risk stemming from uncertainty about the future economy or polity—such as breakdowns of the market or the government, changes in relative prices of labor and capital, or a deterioration in the position of a particular country—by diversifying across types of management (public and private), sources of finance (from labor and capital), and investment strategies (equities and bonds, domestic and international). Risk diversification is especially important given the long time periods and great uncertainty involved. Whatever unpredictable disasters occur in the future, as they surely will, this diversified system is most likely to continue providing protection based on the old adage "don't put all your eggs in one basket." (See World Bank 1994 for more details. For a quantification of the welfare gains from diversification, see Pujol 1996).

How Have Countries Reformed?

During the past decade, and particularly during the past five years, several countries have adopted variations on this multipillar system. The three major variations are the Latin American (individual account) model, the Organisation for Economic Co-operation and Development (OECD) (employer-sponsored) model and the Swedish (notional defined contribution) model. Experience with these variations shows that pension reform is possible, even in democracies, but that it takes somewhat different forms as a result of differing initial conditions and political economies.

For example, different conditions led Argentina to choose a relatively large public pillar, whereas Peru decided against a public pillar. Chile and Australia chose a much larger private pillar than Mexico and Argentina. The United Kingdom and Switzerland built on a history of employer-sponsored plans; Australia and Denmark built upon widespread, union-negotiated plans, and Sweden and Italy adopted a defined contribution plan that remains largely pay-as-you-go.

One of the most important initial conditions that influences the shape of the reform is the implicit pension debt—the present value of the pensions that are owed to current pensioners and to workers according to their years of participation in the old system. This debt is inherent in pay-as-you-go systems, where workers expect to get a specified pension in return for their contributions. But assets are not accumulated to cover this debt; instead, the obligation is covered by implicit government IOUs. In many countries, the implicit debt exceeds the country's conventional explicit debt (backed by government bonds) and in some cases exceeds 200 percent of GDP (table 1). It is especially large in countries with high coverage, generous benefits and older populations. Although this debt is not always legally binding, it tends to be socially and politically binding; governments cannot easily renege on these obligations. Countries that do not want to make their debt transparent frequently shy away from a shift to a funded pillar because it makes at least part of the implicit debt explicit. Most developing countries have little pension debt because of their low

Table 1. *Implicit Pension Debt during the Early 1990s*

Country	*Implicit pension debt as a percentage of gross annual product*	*Percentage of population over 60*
Senegal	27	4.3
Mali	26	4.9
Burkina Faso	15	5.0
Venezuela	30	5.6
Peru	37	5.8
Cameroon	44	5.8
Congo	30	6.1
Brazil	187	6.7
Turkey	72	7.1
Albania	67	8.1
China	63	8.9
Uruguay	214	16.4
Croatia	350	17.8
Ukraine	141	18.7
Hungary	213	19.3

Note: Assuming a 4 percent discount rate.
Source: Kane and Palacios (1996); World Bank data for Albania, Burkina Faso, Congo, and Mali.

coverage rates and are therefore in the enviable position of being able to change to a partially funded system before the debt becomes unmanageable.

The Latin American versus the OECD Models

Chile pioneered the Latin American model in 1980, and its initial success there led Argentina, Bolivia, Colombia, Mexico, Peru, and Uruguay to adopt similar plans in the 1990s. Hungary and Kazakhstan were the first countries outside the region to adopt the model, and it is one of three options proposed by the Social Security Advisory Committee in the United States. In this model, each worker chooses the investment managers of his or her own individual defined contribution retirement account.

By comparison, the OECD model builds on existing employer-sponsored pension plans. These plans simply became mandatory instead of voluntary in Australia, Denmark, Switzerland, (and de facto by collective bargaining in the Netherlands); in the United Kingdom, employer-sponsored plans became an attractive optional alternative to the state plan. Under this model the employer or a combination of employer and union trustees chooses the investment manager for each company or occupational group. These plans thus benefit from economies of scale and financial expertise and possibly from lower marketing costs (although this has yet to be proven). The OECD model, however, introduces a principal-agent problem; that is, the employer or union representative selects the investment manager, but the workers bear the risk. The choice may not be in the worker's best interest and may not maximize net returns. For this reason, workers in OECD model plans may ultimately demand more individual choice.

For example, Australia is now permitting workers to put their retirement savings into special bank accounts. In the United Kingdom employers were initially permitted to opt out of the state earnings-related plan, but subsequently workers were given the right to opt out of their employer's plan in favor of their own personal retirement plan. Unscrupulous insurance company salesmen then persuaded workers to purchase individual annuity plans, when, in fact, they would have been better off staying in the employer's defined benefit plan. The incident, which led to lawsuits and a government inquiry, illustrates the point that worker choice makes consumer information imperative and opens the door to the probability that some mistakes will be made (Johnson, forthcoming; Whitehouse 1998).

In most of the OECD model countries cited above—unlike those in Latin America—a modest public pillar with a small pension debt and little or no payroll tax financing was already in place when the new system was adopted. Thus they could simply retain it and build the second pillar alongside the first. They had no trouble financing the transition because accrued rights were small and the contributions to the second pillar were added on, rather than being diverted from the first pillar.

For example, Australia had a means- and asset-tested first pillar, financed out of general revenues, to which it simply added a mandatory, employer-based, funded pillar financed by payroll contributions. General revenues also financed Denmark's flat benefit in the public pillar, now being reduced, while an occupational funded pillar is added. In the United Kingdom the state earnings-related pension had been initiated just a few years before Margaret Thatcher decided to end it by encouraging employers and workers to opt out; the accumulated rights were still very small. In Switzerland employer-run plans that already existed in many firms became mandatory alongside the public pillar.

In contrast, bloated public pillars in the Latin American countries meant that the first pillar had to be downsized and redesigned to create space for the second pillar. When a worker switched to the new system, he was given credit for his past service under the old system while part of his future contributions were diverted to the new second pillar. These countries had to find the money to continue paying the promised benefits to current pensioners and older workers (the implicit pension debt) under the old system, while part of the payroll tax flowing in was diverted to funded individual accounts—a problem that has become known as "financing the transition." Most countries that reform in the future will have to solve this problem.[1]

The Swedish Model: Notional Defined Contribution Plans

Many countries with large public pillars and implicit pension debts have found it exceedingly difficult to make the transition to a partially funded system with a mandatory private pillar, in part because of the financing problem, but also because of the political interests associated with existing institutions. This explains the third group of reforming countries—those that feature notional defined contribution plans. In this plan the worker has an individual account that is credited with his contributions plus interest. The accumulation is notional, however, rather than actual, since the money paid in by workers is immediately paid out to pensioners rather than being invested; in fact, the system remains pay-as-you-go. Upon retirement, the notional accumulation is converted into a real annuity, supposedly on actuarially fair terms. Thus the notional defined contribution plan is essentially a reformed pay-as-you-go pillar (sometimes accompanied by a second, funded pillar). The problem of transition costs is substantially avoided.

Sweden developed this system, although it has not yet been implemented there. Shortly after Sweden acted, Italy adopted the system, but with a long transition period. In both cases, the first pillar is to be converted into a notional defined contribution plan, buttressed by a redistributive guaranteed pension. The system is also being implemented by Latvia, which hopes to save enough money from reducing

asion and early retirement eventually to start a funded pillar. Poland plans a new stem with a notional defined contribution first pillar and a funded second pillar. utside of Europe, China has a notional defined contribution system, de facto. In inciple China wants to start a second pillar made up of funded individual ac- unts, but many cities have been unable thus far to finance the transition, so the dividual accounts remain largely notional.

The notional defined contribution system was designed to capture some of the lvantages of linking benefits closely to contributions within each cohort. Most nportant, it reduces idiosyncratic intracohort inequities and labor market distor- ons, including incentives to evade—providing the notional interest rate is close to e market interest rate. For example, early and late years of contributions receive the me rate of return, and workers with flat age-earnings profits receive the same rate return as those workers with steep profiles; this equality is not true of most defined nefit plans. In addition linking benefits to contributions makes the system more stainable and avoids the selection problem that occurs when low return people ade but high return people stay in the system.

Furthermore, the notional defined contribution system discourages early retire- ent because workers automatically receive lower benefits if they retire early; and e costs of early retirement are internalized rather than being passed on to others. or the same reason, it automatically induces workers to retire later as longevity creases, thereby avoiding the difficult political decision to raise the retirement age, step that is periodically necessary in defined benefit plans.

The notional defined contribution plan is not inherently redistributive, how- er, so it does not accomplish the first pillar task of protecting low-wage earners. or this purpose a redistributive "O pillar" must be added that will guarantee a inimum pension to workers whose own defined contribution pension falls be- w a specified minimum. Such a guarantee assures a minimum pension even if e employee contributed very little. If the redistributive O pillar is large, it may erode the link between contributions and benefits. In Sweden, for example, the gh level of the guaranteed minimum pension in the O pillar makes the defined ntribution component irrelevant for the majority of workers—especially women, ho are more likely to have spent only a part of their adult lives in the labor rce.

A bigger failing is that the notional system does not capture the benefits of fund- g, because there are no funds. That is, the system serves as the first (pay-as-you-go) llar and crowds out the opportunity for a large funded pillar. Intergenerational ansfers remain, saving is not augmented, and financial markets do not develop. lost important, as the dependency rate increases, the contribution rate would have increase to keep the system solvent in the absence of prefunding. These younger horts may have to "save" a much larger amount for their old age than is optimal

for them in order to cover benefits promised to older cohorts. In that case, the incentives for evasion and escape to the informal sector would be strong.

Sweden plans to build a buffer fund to reduce the need for large tax increases for the first pillar as its population ages. But this buffer fund will be a publicly managed overlay, because the individual accounts remain notional. This situation raises all the problems, summarized earlier, concerning political manipulation and poor allocation of publicly managed funds. The pillar will be supplemented by a small second pillar funded at 2.5 percent of payroll.

How are the notional interest rate and the conversion rate of notional capital into annuities determined? If the notional interest rate is higher than the market rate, it will be a costly guarantee for the government to fulfill. If it is less than the market rate, the contribution is more likely to be regarded as a tax, so labor market distortions are likely to reappear, and pressures may arise for an increase. Typically the interest rate is set equal to some exogenous rate to insulate it from such political manipulation; nevertheless, the possibility remains that a future government will discard this connection and arbitrarily change the rate. Most commonly thus far, the notional rate has been tied to the growth in the per capita wage, or the covered wage bill—supposedly an equilibrating device. If the wage bill increases, so too do contributions, and therefore the ability to impute interest will be high. This means, however, that when the working-age cohort is large and growing (for example, the baby boomers), the imputed interest rate is high, and the pension debt increases rapidly. But when the working age cohort declines (generation X), so too does the notional interest rate. This generation must then pay a high contribution rate to cover the pension debt and will receive a low notional interest rate— fertile grounds for evasion and questionable from the viewpoint of intergenerational equity. Thus using wage bill growth as the notional interest rate does not appear to be an equilibrating device (see Schwarz and Valdes, forthcoming).

The conversion factor into annuities supposedly depends on expected longevity upon retirement. Because the process is notional, however, it too is highly subject to political manipulation. For example, the government can decide to grant notional credit for noncontributing years (a common problem in old pay-as-you-go defined benefit systems), it can impute a low or a high future interest rate into the calculation, and it can fail to adjust the conversion factor when life expectancy increases. In the absence of market discipline, implicit taxes or subsidies can creep in that interfere with the labor market efficiency effects of the new system. Because the government sets both the conversion factor and the interest rate, the notional defined contribution may be thought of as a pay-as-you-go defined benefit in which the benefit is defined in a new way.

In sum, the notional defined contribution system is attractive to countries that have very large implicit pension debts, especially those that are unwilling to incur an explicit fiscal deficit to pay off these obligations. It may be a politically convenient way to

reduce benefits in inflated programs and to equalize the retirement age for men and women (as in Latvia). In such cases it may lay the groundwork for savings that eventually enable the growth of a funded second pillar; but until that happens, it should be recognized as a reform of the first pillar rather than as an introduction of a multipillar system. (For details on the Swedish reform, see Sunden 1998; on Latvia, see Fox 1998; on Poland, see Rutkowsky and others 1997; and on Italy, see Hamann 1997).

How Have Countries Financed the Transition?

When countries with a large pay-as-you-go pension debt shift to a multipillar system that includes a funded component, some of the contribution usually is shifted to individual accounts. Some other revenue source must then be found to cover the resulting financing gap between the remaining revenues and the expenditures needed to pay retirees. The only countries to have experienced this problem are those that have followed the Latin American model. How did these countries finance the transition? Three basic methods were used: reducing the value of the pension debt and the financing gap, finding alternative revenue sources to pay it off, and, finally, resorting to the general borrowing and taxation powers of the treasury.

Reducing the Implicit Debt and the Financing Gap

The implicit social security debt and the gap between payroll taxes and expenditures can be reduced in several ways. First a country can take certain steps before the transition, such as downsizing benefits provided under the old system, raising the retirement age and the penalties for early retirement, tightening eligibility for disability benefits, and changing the indexation method to price indexation, so the outstanding debt, whether implicit or explicit, will be smaller. Argentina, Chile, and Uruguay followed this strategy, which may be indispensable to a good pension reform. It cuts the benefits that must be paid to those who stay in the old system, as well as the compensation owed to those who switch to the new system, and it increases the probability that workers will switch. Otherwise, there is a risk that the government will pay excessive amounts for benefits that never should have been promised in the first place, and it will be more difficult than before to escape from these promises.

Second, the government can acknowledge the value of the pension earned thus far by issuing a recognition bond (as in Chile) or a promise of a compensatory pension (as in Argentina) to each worker who switches to the new system. This step postpones the day when cash will be needed, because the recognition bond cannot be cashed until the worker retires, and the compensatory pension is gradually paid off over the entire retirement period of the worker. Besides extending the pay-off period, the issuance of the recognition bond provides another opportunity to reduce

the debt. A legally binding piece of paper, the bond gives the worker greater certair that the pension debt will eventually be repaid, and in return for reducing unc tainty, the government can downsize the face value of or interest rate on the bond in Peru). The face value can be further reduced if workers have more faith in the n system than in the old one; they (especially young workers) will then be willing switch even with little compensation for their past service. By choosing the mi mum terms that are needed to convince the desired number of workers to switch government can substantially downsize the recognized debt and save on its tran tion costs (as in Hungary).

Third, a government can keep some workers—and their contributions— in t old system. This may be accomplished by excluding some workers, such as the mi tary or the police, from the new system (as in Chile). Argentina gave all worker choice but made the new system attractive mainly to young workers. Colombia o erates the old system side by side with the new one, and workers are permitted switch back and forth. In Uruguay the new funded pillar is compulsory only for ri and young workers and is voluntary for others. The financing gap is reduced becau those employees who remain in the old system continue to contribute to it. I serious danger with this option is that, in an effort to solve a short-run cash-fl problem, these countries have increased their long-term implicit debt by keepi participants in a financially unsound pay-as-you-go system; this solution may tu out to be unsustainable.

Fourth, a government can retain a large pay-as-you-go component in the ne system, so that some revenues continue to flow into the public pillar. Argenti offers a moderate-sized flat benefit in its new public pillar rather than the narrow minimum pension guarantee used in Chile. In Argentina about 60 percent of t total contribution flows into the public pillar. In addition, workers can choose b tween a funded and a pay-as-you-go option for the second pillar. The inflow funds to the first pillar and the pay-as-you-go second pillar help pay current pensio ers and, eventually, the compensatory pension. But if either pillar offers benefits th are too generous (actuarially unsound), the reform will not be sustainable in the lo run—a danger that Argentina faces.

Finding Alternative Revenue Sources

Governments can also pursue policies to offset the revenue gap. One way is to use existing treasury surplus to pay off part of the pension debt. Chile took this path, bu most countries are burdened with fiscal deficits rather than surpluses. Alternativel countries that have a surplus in the social security system can use it to pay off part the debt. The Latin systems generally did not have a surplus, but the U.S. soci security trust fund could be used in this way, if the United States were to make transition. In cases where public enterprises are being privatized, some of the pro

ceeds can be used to pay off the pension debt—a cancellation of long-term assets against long-term liabilities. Peru followed this strategy, Poland is considering it, and Bolivia is also using privatization assets for pension reform.

Measures to reduce evasion and increase coverage will increase system revenues. Although Argentina's plan incorporated such measures, no reduction in evasion has yet materialized. China is considering financing the transition by bringing all workers in township and village enterprises (a rapidly growing group) into the new partially funded system. This expansion of coverage would help to pay off the accumulated pension debt but simultaneously would create new debt to cover the newly enrolled workers who will eventually demand their pensions, and may lead to evasion in the interim. Thus, this strategy produces short-term revenues but runs the risk of undermining the long-run sustainability of the plan and its credibility unless the payroll tax and promised benefits are low.

Using General Borrowing and Taxation

General treasury debt can be used to cover the remaining cash gap in the short run. Because money is fungible, it is not clear to what extent resources for pension reform have come from debt as opposed to other general revenue sources, but government borrowing has usually increased in the early years of reform. In countries with a large implicit pension debt, the use of temporary debt finance is almost inevitable to mitigate the heavy double burden of taxation on the transition generation of workers. Some of this debt may be sold to the pension funds in the new second pillar; government debt and bank deposits have been the largest initial investments of the new pension funds. An important proviso is that government bond sales should be open, transparent, and carry the market interest rate. Pension funds should not be compelled to purchase government bonds. All Latin American countries limit pension funds' overseas investments, however, which virtually ensures that they will have large investments in domestic government bonds.

Is this temporary debt finance problematic? Financial markets might react negatively if they were not previously aware of the size of the implicit pension debt, or if they believed the obligation to repay it was "soft" and has now become "hard," and if either of these beliefs increases the expected default risk on regular bonds. Two pieces of evidence suggest that, so far, the financial market response has been positive. First, the International Monetary Fund recently adopted the position that debt finance earmarked for a pension transition should be allowed beyond the permissible ceiling for other debt, because it is a swap of implicit for explicit debt in the short run and is intended to reduce the overall debt and will thus improve fiscal solvency in the long run. Second, for much the same reason, Hungary's credit rating from Moody's improved after it adopted its pension reform, even though the reform entailed an increase in the explicit debt.

Eventually, the debt should be paid off through taxation. Otherwise there will be no increase in national savings (additional private saving will be offset by additional public dissaving if the implicit debt is simply changed to an on-going explicit debt). The redemption of the debt through tax revenues can be spread over a long period of time—but the longer the payoff the slower the benefits of increased national saving for productive investment. It has been estimated that if half the current pay-as-you-go system in the United States were converted to a funded system, it would take 70 years to pay off the financing gap with a payroll tax rate of about 1.5 percent (Gramlich 1996); it would take roughly the same amount of time in China (Friedman and others 1996).

How Large Are the Efficiency and Growth Effects of Alternative Systems?

The chief theoretical argument for the recommended multipillar system is that it will have a positive effect on efficiency and growth because the old system introduced—or failed to remove—distortions that the reforms will eliminate. A second argument is that the multipillar approach will enhance the financial sustainability of the old-age system and thereby provide better protection for the elderly in the long run. A third argument is that it will improve intergenerational equity.

Efficiency and growth effects are notoriously difficult to quantify and prove, in part because relatively little experience and data are available and in part because, even with the data, it would be difficult to build models that capture all the complex dynamic interactions; that is, it is difficult to specify the counterfactual. Pension reform has several different potential efficiency effects; usually studies focus on one of these while ignoring or holding the others constant. For example, general equilibrium models that analyze labor supply effects often assume perfect capital markets and thereby limit the predicted increases in savings, and vice versa. In this section I summarize the limited empirical research that has been done on these topics, concentrating on the simulated effects in countries that have been considering structural reforms and econometric estimation of the actual effects in Chile, the country that has the longest track record with a reformed system. In general, the beneficial labor market effects come from shifting to a defined contribution system from a defined benefit plan; the beneficial impact on savings comes from shifting to a funded old-age security plan from a pay-as-you-go system; and the financial market impact comes from managing these funds privately.

First, a brief comment on the distinction between efficiency and growth. Greater efficiency, for example, due to a reduction in labor market distortions, increases the

level of output. If some of the increased output is plowed back into investment, as would often be the case, growth also increases. Growth can also be increased without an increase in efficiency. For example, an increase in savings (and consequently growth) may simply indicate an intergenerational or life cycle redistribution that does not increase efficiency because it does not make (or have the potential to make) everyone better off. But such an increase enhances efficiency if the initial rate of saving was suboptimal because of public or private myopia or because of a tax wedge between private and social returns to investment. Both of these conditions are usually alleged as a justification for mandatory retirement saving plans, in which case they would expand both efficiency and growth.

Avoiding Labor Market Distortions

One problem in pay-as-you-go defined benefit systems is the possibility that the high payroll tax will lead to labor market inefficiencies (stemming from distorted decisions about labor force participation, age of retirement, hours worked, choice of job and location, degree of effort, form of compensation, and so on), whereas the contribution in a defined contribution system may be regarded as saving rather than as a tax. Only fragmentary evidence is available about the effect of pension reform on most of these actions. For example, Wise (1997) shows that the labor force participation rate of older men is highly sensitive to the implicit social security tax on labor, stemming from the absence of actuarial penalties on early retirement—the loss of generous defined benefits during years when they continue working induces most workers to stop working before they reach age 60. Countries that have a larger actuarial adjustment in their systems, hence a lower implicit tax on labor, have higher labor force participation rates of older men. Funded defined contribution plans automatically build in this actuarial adjustment, so by extension they should deter early retirement and its negative impact on GDP and the financial solvency of the scheme.

The distortionary labor market effects of traditional systems may be larger in developing countries because escape to the informal sector is easier there, both for workers and their employers. Productivity in the informal sector may be lower because firms have less access to product and credit markets or because technological change is embodied in capital in the formal sector and has an external effect on labor productivity throughout the economy (as discussed in the endogenous growth literature). In addition, regulations that set a minimum wage and other benefits in the covered sector may lead to a wedge between wages and productivity in the formal versus the informal sectors. In simulations for a representative economy, Corsetti and Schmidt-Hebbel (1997) show that a payroll tax rate of 20 percent could cause a massive (47 percent) shift to the informal sector, thereby reducing economywide

growth by more than 1 percent annually. In many Latin American countries the informal sector and small firms in the quasi-informal sector do indeed absorb more than half of the labor force (ILO 1996). Although many other forces are at work, a shift to a defined contribution system, where benefits are closely linked to contributions, might reduce these incentives for informality.

What light does Chile's experience throw on this issue? Between 1980 and 1990, when the average share of informal employment in Latin America increased from 26 percent to 31 percent, it dropped from 36 percent to 31 percent in Chile. Unemployment in Chile fell and wages rose. Edwards (1997) shows that, given reasonable assumptions about the elasticity of labor demand in the two sectors, the pension reform was responsible for a decline of between 2.2 and 3.6 percent in unemployment and an increase of 5 to 8 percent in wages.

In evaluating these numbers and their applicability to other countries, it is important to realize that a shift to defined contribution may not always have this salutary effect. For example, myopic workers may continue to evade contributions because they will not have access to their mandatory savings for many years. In periods when investment returns are low, workers may be especially tempted to evade, preferring to consume or to invest in education, housing, or consumer durables. In Chile returns have been high (more than 12 percent real during the first 15 years), encouraging compliance of most workers. If the payroll tax for pensions is only a part of the total payroll tax, the incentive to escape to the informal sector may remain strong; again, this phenomenon did not apply in Chile, where the total payroll tax was relatively small. Indeed, preliminary evidence from Argentina suggests that evasion has not declined since the new system was established (Valdes-Prieto, forthcoming). In contrast, Chamorro (1992) and Schmidt-Hebbel (1996) found that only 5 percent of potential contributors in Chile had dropped out of the system. (Chile does not even attempt to cover the self-employed, who make up the largest share of evaders in other countries.) It is difficult to be conclusive about this, because it is hard to separate evasion from normal labor force withdrawals and exogenous shifts into self-employment.

Escape to the informal sector under a funded defined contribution plan does not have the same negative effects on the system's sustainability as it does under a pay-as-you-go defined benefit plan, because the costs are borne by the worker in the form of lower benefits rather than being passed on to others in the form of a higher contribution rate. This is a big plus. Nevertheless, it still creates the same problem for labor allocation and productivity and an even greater problem for the workers who may not have an adequate pension and may become a charge on the public treasury when they grow old. So although the initial evidence from Chile is encouraging, it is important to analyze the data on evasion, wages, and employment carefully to determine whether these results are robust and generalizable.

Increased National Saving

A major rationale for fully funded pension plans is that they increase long-term national saving, with positive effects on growth and efficiency. Such saving is important because most savings stay in the country of origin and most of a country's productive investment comes from its own saving, despite the development of global capital markets.

When a country without a prior pay-as-you-go system institutes a multipillar system, consumption will decrease and saving will increase if the mandatory saving rate exceeds the voluntary rate. When a country with an existing pay-as-you-go system replaces it with a multipillar system, national saving increases if benefits are cut or taxes are increased, usually to cover transition costs. In both cases, putting part of the contribution into the worker's own mandatory saving account may be more politically acceptable and less economically distortionary than increasing saving through high taxes that go into the general treasury.

But this increase might not materialize. Mandatory saving may not increase total private saving if individuals find ways to offset required saving by reducing voluntary saving or accumulated assets. In that case, capital may increase in the mandatory pillar but decline in the voluntary pillar. With perfect capital markets, private saving will not increase at all, because people will simply borrow against their mandatory pension saving. A positive saving effect ultimately depends on the assumptions that voluntary long-term saving and assets are small and that borrowing opportunities are limited for substantial groups within the population. The low asset condition probably holds for most slow-growing economies, the limited borrowing condition for most developing countries, and both conditions for low-income households in most countries.

Public saving matters as well as private saving. On one hand, pension reform may educe public dissaving as governments no longer need to borrow to cover escalating ension costs, but on the other hand, it may increase public dissaving if the build-up pension reserves relaxes fiscal discipline and makes it easier for governments to n large deficits. If the transition is fully financed by borrowing, government dissaving ll offset private saving, and the expected increase in national saving will simply not cur. But if it is financed through taxes or cutbacks in other government expenditures, public saving increases national saving. Estimating the impact on public sav-g therefore requires modeling government behavior—how governments will behave after pension reform and how they might have behaved in the absence of reform.

Several simulations have projected the impact on saving of a shift to a fully funded heme. Not surprisingly, the results turn out to be highly dependent on the as-mptions, especially the assumptions about the crowding out of voluntary saving d the method of financing the transition. Underlining the importance of the former, mulations of a representative economy indicate that a tax-financed transition to a lly funded system in the presence of credit constraints on consumers (implying

low crowd-out) will increase output by 22 percent and welfare by 16 percent in the long run, while the gain is only 2 percent without credit constraints (Cifuentes and Valdes-Prieto 1997).

In planning its new funded pillar, Australia assumed that half of the mandatory saving would be offset by reduced voluntary saving (the crowding-out effect) in the case of newly covered workers and even more for workers who already were covered by voluntary occupational plans. This assumption implied that when the contribution rate reached 12 percent, national saving would rise by 1.5 percent of GDP, almost doubling the current net national saving rate, which is 2.2 percent of GDP. (The gross national saving rate is about 15 percent of GDP.) Australia did not have to borrow to pay off a pension debt because the second pillar was an add-on rather than a diversion of previous contributions. Although the tax deductibility of contributions was initially projected to cause some government dissaving, in the long run the decreased burden on the means-tested public pension is expected to reduce government dissaving. One of the main effects of the reform may be to shift the allocation of private saving away from home ownership, which is now the predominant investment, because of tax inducements, and toward other, more productive, investments (Bateman and Piggott, forthcoming).

In his simulations for Mexico, Ayala (1996) assumes a 30–40 percent rate of crowding out. If the transition is tax-financed, or if it is debt-financed and Ricardian equivalence holds (so that private saving goes up to offset public dissaving), total saving rises by 0.4 percent to 2.1 percent of GDP, a magnitude similar to that expected in Australia. If the transition is debt-financed and Ricardian equivalence does not hold, the impact on total saving is much smaller, even negative in some years although positive overall during the next 30 years.

Only Chile has had a mandatory saving plan long enough to permit an estimate of its effects. Data from Chile are problematic, and the savings ratio is erratic, complicating the analysis and making the results highly sensitive to the starting date for comparisons. According to Corsetti and Schmidt-Hebbel (1997), private sector saving as a percentage of GDP increased from almost zero in 1979–81 to 17 percent in 1990–92, while private consumption decreased commensurately. Their reduced-form two-stage-least-squares regressions attribute half of the decline in the private consumption ratio to the growth of Chile's funded pension plans and correlated developments, such as capital market deepening. Time series regression analyses by Haindl (1996) indicate that pension reform accounts for 6.6 of the 9.9 percentage-point increase in the national saving rate in Chile (from 16.7 percent of GDP in 1976–80 to 26.6 percent in 1990–94). Of the 6.6 point increase, 3.1 points were attributable to the direct impact of pension saving; the remaining 3.5 points were attributable to the financial market deepening caused by the pension fund (4.2 percentage points) offset slightly by a crowding-out effect caused by borrowing constraints (0.7 points). Using an error correction model, Morande (1996) also finds a significant positive

fect of a pension fund dummy on private saving from 1960 to 1995. He speculate iat the financial market deepening caused by pension reform may have made vol ntary saving less likely to be crowded out by, and therefore less sensitive to fluctua ons in, foreign saving, *thus making the country's supply of investible resources les* ependent on foreign capital.

Agosin, Crespi, and Letelier (1996) are more skeptical; they find that increase ivate saving resulted from an increase in corporate saving, from 6 percent of GDP i 978–85 to 23 percent in 1994—a response, they believe, to the lack of foreig edit and the privatization of public enterprises. (Of course, privatization was itsel cilitated by the pension reform, illustrating the complex interactions among thes iriables.) Voluntary household saving was negative (about 4 percent of GDP through it this entire period), indicating consumer dissaving or borrowing. Forced savin *irough the new pension system gradually grew to almost 4 percent of GDP*, how er, and was not offset by greater voluntary dissaving (presumably because credi onstraints had already been exhausted). This 4 percent magnitude is roughly con stent with the findings of Bosworth and Marfan (1994) that pension reform in eased saving by 3 percent of GDP. The risk remains that the growth of consume edit, possibly fueled by the pension reform, could increase consumer dissaving an offset some of these gains in the future (Holzmann 1996).

Instead of focusing on enhanced private saving, other studies emphasize the im pact of pension reform on public saving and dissaving. Chile financed a pensio *transition in part through deficit finance*, which decreased national saving. The fisca costs of the transition may have canceled out the positive effect on private savin initially (Agosin, Crespi, and Letelier 1996). Observing that the pension-related deficit of the government (payments to pensioners from the old system plus redemptions o recognition bonds for new pensioners who had switched) were larger than the in flows to the new pension funds until 1989, Holzmann concludes that during th 1980s the new pension system had a negative effect on national saving. He appears however, to overlook the fact that redeemed recognition bonds became part of pri vate pension saving and were not immediately consumed. Correcting for this poin alone generates a positive savings effect as early as 1985.

More important, a simple accounting exercise neglects the disciplining effect tha pension reform might have had on other government taxes and expenditures. Chil ran an increasing surplus during this period, possibly to help cover the transitio costs. Since 1987 the consolidated government budget has been in surplus, whic quickly exceeded 5 percent of GDP. In addition Chile accumulated a large budgetar surplus in preparation for the reform, thereby reducing its need for deficit finance How large the current or past surplus would have been otherwise is unknown, but to the degree that the pension reform was financed by increasing general taxes, cutting other public spending or accumulating a prior surplus, transition costs did not de crease public saving. Moreover, the transition costs are short run, while the increase

private saving may persist in the long run. As a result of all these factors, total n
tional saving in Chile is currently much higher than it was before reform.

Given the high correlation between pension reform and other policies that a
often simultaneous, the controversy surrounding the determinants of private savir
(for example, which variables are endogenous?), and the even greater uncertain
about the determinants of public saving (what is the counterfactual?), all these econ
metric and simulation results are highly sensitive to the model's specifications. Ne
ertheless, preliminary evidence indicates that pension reform can have beneficial e
fects on long-term national saving, especially if it is accompanied by broader polici
designed to constrain consumer and government borrowing.

Financial Market Development

One reason for favoring private management of pension funds is that it will develo
a set of financial institutions—investment managers, insurance companies, an
banks—that are essential for economic development. A funded pillar, if competi
tively managed and well regulated, can enable the financial market to grow in safet
size, depth, and complexity. In developing countries, where private saving is alread
high, one of the main effects of a funded pillar may be to shift savings out of land an
jewelry and into long-term financial market investments that are better for the broad
economy.

Even in Australia the financial market is expected to grow as a result of the man
datory second pillar. As noted earlier, some private saving may be redirected out o
owner-occupied housing into the financial markets. Insurance companies are ex
panding, developing new products, including annuities, to meet the anticipated de
mand stemming from pension funds (Bateman and Piggott, forthcoming). In Swit
zerland too, the growth of the life insurance industry, investment companies, an
mutual funds have been spurred by mandatory funded pension plans. And corporat
governance has been gradually changing, as institutional investors have demanded
disclosure and better performance (Hepp, forthcoming). All these changes enhanc
efficiency.

But the strongest evidence for this expected growth effect comes from Chile. During
the five years preceding the adoption of its new system, Chile prepared the ground
work by organizing a primary market for treasury bonds, reforming the laws govern
ing mutual funds, corporations and securities, privatizing banks, authorizing a price
indexed mortgage bond market, and liberalizing the provision of insurance and
reinsurance (Valdes-Prieto, forthcoming). Once the system was introduced, this pro
cess continued; financial markets became more liquid as stock market trades in
creased; demand was created for the equities of newly privatized state enterprises
information disclosure and credit-rating institutions developed; the variety of finan
cial instruments grew, and asset pricing improved. In several of the studies summa

rized above, financial market deepening associated with pension reform was given credit for the observed increase in private saving. Econometric analysis suggests that financial market efficiency induced by the pension reform (and related factors) increased total factor productivity 1 percent a year, or half of the increase in total factor productivity (Holzmann 1996).

In sum, a small but growing body of empirical evidence indicates that pension reform has produced positive efficiency and growth effects. That is, the impact on saving, productivity, output, and welfare may be high relative to exogenous sources of growth and other policies available to increase growth.

Several caveats are essential in interpreting this evidence. First, because specifying the counterfactual is difficult, these results are highly sensitive to the assumptions that are made. In particular, the econometric analyses for Chile are subject to omitted variable bias, and the simulation results depend heavily on assumptions about crowding out, transition costs, and rates of return. Second, the growth impact also depends on key policy decisions in setting up the new system, such as the question of how high the required contribution rate will be, the proportion of the multipillar system to be funded and defined contribution, and how the transition will be financed. Debt finance may be necessary for political purposes, but some degree of tax finance is necessary to meet the economic objectives, and of course some taxes have better efficiency properties than others. Third, it is important to remember that, even if it claims to use a general equilibrium model, each study typically deals with only one possible source of growth, so many of these results are partially additive—that is, the total growth effect is the sum of the separate effects on labor market distortions, early retirement, escape to the informal sector, capital accumulation, financial market development, and other sources of growth. So if each separate effect increases GDP in amounts ranging from 1 to 10 percent, their sum may increase GDP much more.

New Problems and Issues for Further Research

Although many efficiency gains seem to have been achieved, the new systems have also created problems that must be solved and related research that needs to be done. The problems involve high administrative costs, financial market distortions, and distributional effects.

Administrative Costs

The big advantage of private investment over public is the likelihood that it will produce a better allocation of capital and therefore higher returns for the fund and growth for the economy. Decentralized systems also may charge high administrative

fees, however, partly because of high marketing costs in competitive industries. Costs sometimes produce important side benefits, such as consumer information and increased compliance, but this does not appear to be the case in most countries that have recently reformed.

Preliminary evidence indicates that workers are ill informed and do not make decisions based on fees or investment returns, and that pension funds incur high sales commissions and other marketing costs to attract them. In Chile and other Latin American countries, fees are front-loaded, meaning that workers pay a one-time fee on new contributions rather than an annual fee based on assets. (This system was probably adopted because the new system initially had no assets.) Specifically, this one-time fee is about 2 percent of wages or 15–20 percent of new contributions, and about one-third of this fee is for marketing.

These numbers appear very high. To understand their impact on net returns these one-time charges on contributions must be converted into their equivalents in annual charges on assets, a conversion that depends on how large the assets are relative to the contributions. Obviously, for accounts that have small accumulated assets (young workers with few years of contributions), the one time fee will be high relative to assets. For accounts that have built up substantial assets over the years, however, the fee will be small relative to assets.

Simulations show that if the current fee schedule is maintained, the average Chilean worker who contributes for 40 years will pay the equivalent of less than 1 percent of assets a year. This is approximately the same amount mutual funds charge for voluntary retirement savings accounts in the United States; it is not excessive, from the lifetime point of view, in comparison to a competitive market retail price for individuals. Moreover, it is not excessive in comparison to a less expensive system that produces much lower gross and net returns (such as publicly managed reserves in Singapore and the United States social security trust fund). Competition may lower costs further in the long run.

Nonetheless, this fee structure is an apparent problem in the early years of a new system, when all accounts are small. It is a real problem for workers who will be in the system for only 20 to 30 years, such as workers who were relatively old upon the date of reform; simulations show that these workers pay a much higher lifetime fee as a percentage of assets. It is a problem for transient workers who move in and out of the labor force, such as women, because they may never have a chance to accumulate 40 years of contributions. The higher lifetime fee as a percentage of assets and hence the lower net return received by these groups is a matter of concern on equity grounds in a mandatory system. On both equity and efficiency grounds, it is questionable whether the government should compel all workers, including those who are very risk-averse, to incur these costs with certainty while the benefits are uncertain. Besides the equity consideration is the practical consideration that high costs may lead those groups most affected to evade. Moreover, it would be desirable to find ways to

increase administrative efficiency for all workers to increase their rates of return and replacement rates.

Some analysts believe that administrative costs would be lower under a group plan, and they thus favor choice by the employer or union. Such group plans may be better positioned to benefit from economies of scale in decisionmaking, greater financial expertise, and lower marketing costs (for a discussion of scale economies see James and Palacios 1995; Mitchell 1996a). This is one reason why employers or unions choose the investment manager in OECD model countries. Because employers or union representatives make the investment decision while workers bear the risk, such plans can also open the door to financial abuse and principal-agent problems: employers might choose investment managers or strategies that benefit them even if their choice implies lower returns for their workers.

For example, lower "wholesale" charges appear to be available for large group 401(k) plans in the United States, but not all employers have gone to the effort of obtaining these rates. In Switzerland employers tend to place retirement funds at banks with which they have longstanding financial relationships, without exploring other options carefully (Hepp, forthcoming). One of the worst cases of employer abuse of worker retirement funds was the Maxwell scandal in the United Kingdom; but individual choice also led to a scandal in that same country, as uninformed workers were induced to abandon their employers' plans and purchase financially disastrous policies by unscrupulous insurance company salesmen (Johnson, forthcoming). Basing the second pillar on occupational plans is especially a problem for mobile workers, who may end up with many small costly accounts unless these can be consolidated in one personal account.

Anecdotal evidence about costs and returns to group choice versus individual choice is available, but a careful empirical study has yet to be done. Meanwhile, the principal-agent problem makes it likely that political pressures will develop to give workers the right to opt out of employer pension plans into their own personal retirement savings plan in most mandatory systems; this has already happened in the United Kingdom and Australia.

A third alternative may be desirable in small countries where economies of scale do not allow markets to support many pension companies efficiently, in countries with undeveloped financial markets that want to attract investment expertise and minimize start-up costs, and in countries with low contribution rates to the second pillar. Instead of open entry, the government might auction off operating rights to a limited number of investment companies, among whom workers then choose. The contract could specify the maximum risk, offer a reward for high returns, and choose the winners based on who charges the lowest administrative fees.

The voluntary Thrift Saving Plan for United States federal employees uses a competitive bidding process to choose its money managers, at a total cost of less than 10 basis points (0.1 percent). An auction process was recently used in Bolivia, which as

a result expects to have much lower administrative costs than Chile does. Another alternative is to set a low fee ceiling and open entry to all qualified pension fund managers willing to abide by that limit. Sweden plans to use a variation on this theme for its new second pillar—centralized collection and record-keeping, while workers choose among mutual funds that have reached an agreement on fees with the central agency.

The dangers here are the difficulties in insulating the auction and investment process from political manipulation, corruption, and collusion, and in incorporating incentives for good performance when entry and price are limited. Moreover, these mechanisms may feature less consumer education and service along with lower marketing expenditures. The advantages are that much lower costs, allowing an increase in net rates of return and replacement rates of as much as 20–40 percent, can be achieved if the process is well handled.

To sum up, one could construct a continuum with considerable choice, competition, political insulation, and relatively high administrative costs on one end and limited choice and competition with lower costs on the other end, with each arrangement having different implications for political insulation, rates of return, and other kinds of service. Countries could then choose which mix of costs and benefits they prefer. Thus many additional measures can be and are being considered to economize on costs, and their effects should appear over the next decade. The impact of alternative institutional arrangements on administrative costs in the second (decentralized funded) pillar has heretofore received little attention. Society could certainly benefit from careful analytic and empirical studies in this area.

Financial Market Distortions

Multipillar systems have justifiably been given credit for stimulating the growth of financial markets in middle-income countries and thus promoting economic growth. As these systems have been implemented, however, they have distorted the operations of financial markets in various ways.

This problem results because policymakers want workers to make investment decisions and bear the corresponding risk, but they also want to limit this risk to avoid a disaster. Relatedly, the government must set certain investment constraints and offer guarantees to overcome political opposition to reform. The contradiction here can potentially lead to malfunctioning markets, particularly if the pension funds are relatively large players in the market. Although the risk-reducing benefits of international diversification and diversification into private sector securities is one of the rationales for pension reform, in fact most countries require or strongly encourage domestic investments, with a heavy concentration in government bonds.

This ambivalence can be seen in Chile and several other Latin American countries where pension funds are heavily penalized if they deviate more than 2 percentage

points from the group mean. As a result, funds have been accused of herding behavior, as each tries to look very much like the others. Rather than having a choice of different points on the risk-return frontier, stemming from differing asset allocations—as would be the case in a well-functioning financial market—workers have the much less meaningful choice of companies that provide the same asset allocation and risk-return mix. Because workers are required to invest in one fund instead of diversifying among several and thereby reducing their risk, the lack of meaningful portfolio differences among them means that gains from diversification would be small in any event.

In Mexico all workers are required to enter the new system, but those currently in the labor force may return to the old pay-as-you-go system upon retirement if they will fare better under it. This option was included to acknowledge the "acquired rights" of workers and therefore avoid a legal challenge to the reform efforts. But it creates an obvious moral hazard problem: workers have an incentive to gamble with their pension funds, accepting too much risk because they are substantially protected from loss. The Mexican authorities have avoided the problems by greatly limiting the choice of investment strategies: at least 65 percent of all assets must be invested in government bonds (currently the funds have 99 percent of their assets in government bonds) and international investments are proscribed. Because workers have no real choice of portfolios, moral hazard is avoided; but the flow of pension funds to the financial market and the private sector is also avoided.

Bolivia initially intended to invest most of its revenues from privatization (targeted for pension reform) abroad, to protect it from excessive government borrowing and other country specific risk. To overcome union opposition to the reforms, however, the government had to pay off the implicit debt of the complementary pensions that unions had negotiated in the past. To cover these and other expenditures, the final arrangements decreed that initially almost all of the privatization assets would be invested domestically, in government bonds. In Uruguay, to help cover transition costs, pension funds are required to put at least 80 percent of new assets in special issue government bonds.

Regulations in Switzerland require a 4 percent nominal guaranteed rate of return on their second pillars, thereby leading to a very conservative investment strategy, consisting largely of bonds. Until recently, providers of second pillar pensions for civil servants in the Netherlands faced little competition, again leading to low rates of return that might have been off or at a corner of the risk-return frontier.

These distortions should not be exaggerated, because the guarantees and limits on competition and portfolio diversification are likely to fall through time, as the schemes mature. Chile started with rigid restrictions but has gradually opened up the system to greater diversification, including international investment. Mexico is now considering allowing each pension fund to offer more than one portfolio, together with worker diversification among different portfolios. Another possibility under consid-

eration is to allow pension funds to differentiate their asset allocation strategies and corresponding benchmarks (if available), applying different risk limits depending on the type of portfolio chosen. For example, the Thrift Saving Plan in the United States offers portfolios that concentrate on bonds, stocks, and international investments, with different degrees of risk implied by each. This choice allows workers to pick their preferred point on the risk-return frontier and should help the financial markets to operate better, but it also requires substantial worker education as well as a greater diversity of financial instruments than currently exists in many developing countries.

The Distributional Impact of Pension Reform

Although this paper has focused on the efficiency and growth effects of pension reform, an equally important topic is the impact of reform on equity. Because traditional pension systems are typically both inefficient and inequitable, they offer an opportunity to improve both. Which multipillar systems have actually succeeded in achieving a better distributional outcome is not known, however. Closer examination suggests that the devil is in the details, and some of the results may be surprising.

For example, in Chile's public pillar, workers are eligible for a minimum pension guarantee of about 27 percent of the average wage after 20 years of contributions, meaning that the government tops up the benefits of these workers to the guarantee point if their own accumulation does not suffice. The main beneficiaries here will be lower earners who worked only 20 years, disproportionately females, who have limited labor market attachment, while workers who remain in the formal sector for a full career are unlikely to receive this subsidy. In contrast, in Argentina a flat benefit of about 28 percent of the average wage is paid to all workers who have at least 30 years of contributions (plus an additional 1 percent for every year above 30 up to 45). The main recipients will be workers who spent most of their adult lives in the formal labor sector, and (in sharp contrast to Chile) women are unlikely to qualify. In the United Kingdom, which pays a flat benefit about half the size of Argentina's (as a proportion of the average wage) but does not set a required number of contributory years, the big gainers are people who work few years and live long lives, such as women.

The setup of the second pillar also has distributional consequences. If flat fees per account are permitted, net returns are reduced for low earners more than for high earners. Flat fees were charged by Chilean funds initially, but the unfavorable publicity they encountered was one factor leading them to drop this practice. Some funds in Mexico now use flat fees. If low-income workers tend to choose more risk-averse investment strategies than high-income workers, they will have lower replacement rates in the future. The distributional issue is explored further in a separate paper (James 1997) and certainly merits additional empirical research.

Conclusion

Averting the Old Age Crisis (World Bank 1994) argued that old-age security systems with a large funded defined contribution component, decentralized competitive fund management, and a social safety net, are most likely to promote economic growth, provide acceptable income to the old, and reduce risk by diversification. During the past five years, the move toward multipillar systems has accelerated. With the aging of the global population, it has become increasingly important to choose a reliable and cost-effective method of old-age support. As economic growth slows and financial markets open up, it has become increasingly important to raise productivity through improved incentives in the labor market and through the accumulation of capital that can be allocated to its most efficient uses. To reduce income disparities, it has become increasingly important to provide additional protection to low-income wage-earners who have grown old. A multipillar system that includes a mandatory, publicly managed, tax-financed defined benefit pillar for redistribution, a mandatory, privately managed, funded defined contribution pillar to manage peoples' retirement savings, and a voluntary pillar for people who are willing to pay for more security, has seemed to many countries the most likely way to accomplish these objectives.

Thus several Latin American, OECD, and transition countries have already adopted multipillar systems, and they are under serious consideration in many more. Preliminary evidence from Chile, the only country that has had this system in place long enough for empirical studies to be conducted, supports the existence of a positive growth effect, stemming from increased labor market efficiency, mobilization of long-term saving, and financial market development.

Countries with a large implicit pension debt and an accompanying set of social security institutions are having trouble overcoming political opposition and financing the transition, however. Developing countries are fortunate in that they are at a relatively early stage and can choose a preferred multipillar system almost from the start before these obstacles arise.

Notes

1. Estelle James is lead economist in the Development Economics Research Group at the World Bank. She would like to acknowledge the assistance of Robert Palacios, who gathered the data for figure 3, and Cheikh Kane who collaborated with Mr. Palacios in the preparation of table 1.

2. For further details on the Latin American and OECD reforms see Bateman and Piggott (forthcoming); Cerda and Grandolini (1997), Hepp (forthcoming), Johnson (forthcoming); Mitchell (1996b), Palacios and Rocha (1997); Queisser (1998); Rofman and Bertin (1997); Valdes-Prieto (forthcoming), Von Gersdorff (1997), and Whitehouse (1998). For another summary of structural and piecemeal reforms see Demirgüç-Kunt and Schwarz (1997).

References

The word "processed" describes informally reproduced works that may not be commonly availabl through library systems.

Agosin, Manuel R., Gustavo Crespi, and Leonardo Letelier. 1996. "Explicaciones del Aumento del Ahorro en Chile." Centros de Investigacion Economica, Santiago, Chile: Banco Interamerican de Desarrollo.

Ayala, Ulpiano. 1997. "The Savings Impact of the Mexican Pension Reform." Green Cover Report 16373–MF. Mexico Country Management Unit, vol. 2. World Bank, Washington, D.C. Processed.

Bateman, Hazel, and John Piggott. Forthcoming. "Mandatory Occupational Pensions in Australia *Annals of Public and Cooperative Economics.*

Bosworth, Barry, and Manuel Marfan. 1994. "Saving, Investment, and Economic Growth." In Barr Bosworth, Rudiger Dornbusch, and R. Labán, eds., *The Chilean Economy. Political Lessons an Challenges.* Washington, D.C.: Brookings Institution.

Cerda, Luis, and Gloria Grandolini. 1997. "Mexico: The 1997 Pension Reform." Paper presented at the Economic Development Institute (EDI) conference on Pension Systems: From Crises Reform, November 21–22, Washington, D.C. Processed.

Chamorro, Cerda. 1992. "La Cobertura del Sistema Chilean de Pension." Ph.D. dissertation Universidad Catolica, Santiago, Chile. Processed.

Cifuentes Rodrigo, and Salvador Valdes-Prieto. 1997. "Transitions in the Presence of Credit Con straints." In Salvador Valdes-Prieto, ed., *The Economics of Pensions Principles, Policies, and Inter national Experience.* New York: Cambridge University Press.

Corsetti, Giancarlo, and Klaus Schmidt-Hebbel. 1997. "Pension Reform and Growth." In S. Valde Prieto, ed., *The Economics of Pensions. Principles, Policies and International Experience* New York Cambridge University Press.

Demirgüç-Kunt, Asli, and Anita Schwartz. 1997. "Taking Stock of Pension Reforms around World." Paper presented at EDI conference. Processed.

Edwards, Sebastian. 1997. "Chile: Radical Change Toward a Funded Pension System." Paper pre sented at Kiel Week Conference, University of Kiel, Germany. NBER, Cambridge, MA. Process

Fox, Louise. 1998. "The Latvian Pension Reform." World Bank, Social Protection Group. Was ington, D.C. Processed.

Friedman, Barry, Estelle James, Cheikh Kane, and Monika Queisser. 1996. "How Can China P vide Income Security for Its Rapidly Aging Population?" Policy Research Working Paper 16 World Bank, Policy Research Department, Washington, D.C. Processed.

Gramlich, Edward. 1996. "Different Approaches for Dealing with Social Security." *Journal of E nomic Perspectives* 10(3):55–66.

Haindl, Erik Rondanelli. 1996. "Chilean Pension Fund Reform and Its Impact on Saving Discussion Paper 96-2. University of Miami, International Business Center, Coral Gables, Flor Processed.

Hamann, A. Javier. 1997. "The Reform of the Pension System in Italy." IMF Working Paper 97 Washington, D.C. Processed.

Hepp, Stefan. Forthcoming. "The Swiss Multi-Pillar System." *Annals of Public and Cooperat Economics.*

Holzmann, Robert. 1996. "On Economic Usefulness and Fiscal Requirements of Moving fr Unfunded to Funded Pensions." Working Paper. University of Saarland, Saarbrucken, German Processed.

, (International Labor Organization). 1996. PANORAMA. Geneva.

nes, Estelle. 1997. "Pension Reform: Is There an Efficiency-Equity Tradeoff?" In Nancy Birdsall, Carol Graham, and Richard Sabot, eds., *Beyond Trade-offs: Market Reforms and Equitable Growth in Latin America*. Washington, D.C.: Inter-American Development Bank and Brookings Institution.

nes, Estelle, and Robert Palacios. 1995. "Costs of Administering Public and Private Pension Plans." *Finance and Development* 32:12 15.

inson, Paul. Forthcoming. "Pension Reform in the United Kingdom." *Annals of Public and Cooperative Economics.*

ne, Cheikh, and Robert Palacios. 1996. "The Implicit Pension Debt." *Finance and Development* 33:38-42.

tchell, Olivia. 1996a. "Administrative Costs in Public and Private Pension Systems." Working Paper National Bureau of Economic Research, Cambridge, Mass. Processed.

—. 1996b. "Social Security Reform in Uruguay." Pension Research Council Working Paper 96-20. University of Pennsylvania, Philadelphia., Pa. Processed.

orandé, Felipe G. 1996. "Savings in Chile: What Went Right?" Working Paper 322. Inter-American Development Bank, Washington, D.C. Processed.

lacios, Robert, and Roberto Rocha. 1997 "The Hungarian Pension System in Transition." Paper presented at EDI conference Processed.

jol Thierry 1996 "Some Considerations on Pay-As-You-Go vs. Fully-Funded Schemes." World Bank Washington, D.C. Processed.

eisser, Monica 1998 "The Second Generation Pension Reforms in Latin America." OECD, Paris Processed

fman, Rafael, and H Bertin 1997 "Lessons from Pension Reform: The Argentine Case." EDI conference. Processed

akowsky, Michael, and others 1997 "Security through Diversity. Reform of the Pension System in Poland." Office of the Government Plenipotentiary for Social Security Reform, Warsaw. Processed.

hmidt Hebbel, Klaus. 1996 "Pension Reform, Informal Markets, and Long-Term Income and Welfare" Paper presented at EDI conference. Processed.

hwarz, Anita, and Salvadore Valdés. Forthcoming. "National Accounts: An Analytic Framework." Social Protection Group, World Bank, Washington, D.C. Processed.

nden, Annika. 1998 "The Swedish National Account System." Social Protection Group, World Bank Washington, D.C. Processed.

lés-Prieto, Salvador Forthcoming. "Pension Reform in Latin America: Transition and Implementation Problems." *Annals of Public and Cooperative Economics*

n Gersdorff, Hermann. 1997 "Pension Reform in Bolivia. Innovative Solutions to Common Problems" Policy Research Working Paper 1832. World Bank, Financial Sector Development Department, Washington, D.C. Processed.

itehouse, Edward. 1998. "Pension Reform in Britain." Social Protection Group Discussion Paper 98-10. World Bank, Washington, D.C. Processed.

se, David. 1997. "Retirement across Decades, across Nations." Paper presented at International Health and Retirement Surveys Conference, Amsterdam. Processed.

rld Bank. 1994. *Averting the Old Age Crisis. Policies to Protect the Old and Promote Growth.* A Policy Research Report. New York: Oxford University Press.

The Evolution of Poverty and Inequality in Indian Villages

Rajshri Jayaraman • Peter Lanjouw

... paper examines the evolution of poverty and inequality in rural India by reviewing ...gitudinal village studies. It explores the main forces of economic change—agricultural ...nsification, changing land relations, and occupational diversification—from a wide ...ge of disciplinary perspectives, and it considers the roles of various institutions as con...ts of change. Although most village studies support the survey-based judgment that ...l poverty declined in India during the 1970s and 1980s, they find that progress has ...n slow and irregular and that inequalities within villages have persisted. These con...ued inequalities may constrain both the scope for further poverty reduction from eco...mic growth and the impact of policy interventions.

...ormation on rural living conditions in India is abundant, compared with that for ...st other countries. Sample survey and census data, collected regularly since inde...ndence, have been used to trace the evolution of consumption levels and poverty ...s, demographic and occupational trends, educational levels, and health and nu...tional status. These data have been analyzed at the national as well as district levels ...recent examples, see Drèze and Sen 1995; Datt and Ravallion 1996; Drèze and ...ivasan 1996; and Sen 1996).

...here is also a rich tradition of village studies in India, reflecting a wide range of ...plinary backgrounds and methodologies. These studies bring context and per...ctive to our understanding of Indian rural life, highlighting relationships between ...seholds and their surrounding community and illustrating the roles played by ...age institutions.

...lthough survey and census data have been widely analyzed, there have been only ...w attempts to bring together the findings of the numerous village studies (but see ...ton with Longhurst 1989; Harriss-White 1992; and for comparisons of survey ...d village studies, Bardhan 1989 and Harriss-White 1996). This paper attempts ...h a review, focusing on the subset of Indian village studies that analyze change ...r time. Most of the studies reviewed involved at least one revisit to the village. A

...orld Bank Research Observer, vol. 14, no. 1 (February 1999), pp. 1–30

few record either single, very long-term visits or projects in which the researcher made a particular effort to record change over time. The paper also draws on several studies that take a broader, regional, perspective.

There is no statistical basis for generalizing beyond the village studies reviewed. Although we examine a large number of studies, covering a wide range of locations, we do not have a random sample. The study villages were selected for various specific reasons. If they share one common characteristic, it may be that they are, on average, less remotely located than villages are in general. The data collected, moreover, are dissimilar, limiting ready comparisons. One of the challenges in undertaking our review, in fact, has been to interpret the findings of investigators using varying methodologies and wearing differently tinted spectacles. The degree to which their findings can be extrapolated to rural areas in general is thus ultimately a question of judgment and personal inclination.

Table 1 lists the 35 longitudinal village studies reviewed in this paper. The regional coverage is fairly broad, although it is clear that Tamil Nadu and, to a lesser extent, Uttar Pradesh, are the most heavily represented. Important omissions include Madhya Pradesh, Haryana, Kashmir, Himachal Pradesh, and the northeastern states. In addition, the information available for Orissa, Karnataka, and Kerala is relatively old.

Table 1. *Longitudinal Village Studies Reviewed*

State	*District or region*	*Name of village*	*Period studied*	*Author*
Andhra Pradesh	Mahbubnagar	Aurepalle, Dokur	1975–85	Walker and Ryan (1990)
Andhra Pradesh	Chittoor, Guntur		1950–88	Da Corta (199[illegible])
Bihar	Purnea, Saharsa, Katihar	Five unnamed villages	1971–84	Rodgers (1983, 1984)
Bihar	Purnia	Majra, Barahari	1985–88	Jha (1994)
Gujarat	Bulsar	Gandevigam, Chikhligam	1962–91	Breman (199[illegible])
Karnataka	Mysore	Wangala, Dalena	1954–70	Epstein (197[illegible])
Karnataka	Mysore	Rampura	1948–70	Srinivas (19[illegible]6)
Kerala	Aleppey, Palghat	Two unnamed villages	1976–77	Mencher (1980)
Maharashtra	Vidarbha	Olegao	1920–70	Attwood (19[illegible]9)
Maharashtra	Vidarbha	Palso	1947–87	Gadre, Wahlk[illegible] Galgalikar [illegible]
Maharashtra	Sholapur, Akola	Shirapur, Kalman, Kanzara, Kinkheda	1975–85	Walker and Ryan (1990)
Orissa	Phulbani	Bisipara	1950s	Bailey (195[illegible])
Punjab	Ambala	Shahidpur	1965–78	Leaf (1983)
Rajasthan	Nagaur, Jodhpur	Two unnamed villages	1963–84	Jodha (1989)

Table 1 *(continued)*

State	*District or region*	*Name of village*[a]	*Period studied*	*Author*
Tamil Nadu	Thanjavur	Kumbapettai, Kirippur	1950s–1980s	Gough (1981, 1987)
Tamil Nadu	South Arcot	Iruvelpattu	1916–81	Guhan and Mencher (1983)
Tamil Nadu	Madurai	Gokilapuram	1977–86	Ramachandran (1990); Swaminathan (1988, 1991)
Tamil Nadu	Thiruchirapalli	Rajendram, Povvamani, North Nangavaram, Naganur, Kalladai, Peripatti	1979–80	Athreya, Djurfeldt, and Lindberg (1990)
Tamil Nadu	North Arcot	Randam	1973–84	Harriss (1991)
Tamil Nadu	North Arcot	Eleven unnamed villages	1973–84	Hazell and Ramasamy (1991)
Tamil Nadu	Coimbatore	Three unnamed villages	1981–82	Heyer (1992)
Tamil Nadu	Thiruchirapalli	Poovalur	1987–88	Kapadia (1993)
Tamil Nadu	Tiruvannamalai, Sambuvarayar	Nesal, Vinayagapuram, Veerasambanur	1973–93	Janakarajan (1996)
Uttar Pradesh	Mainpuri	Karimpur	1925–84	Wiser and Wiser (1971); Wadley and Derr (1989)
Uttar Pradesh	Moradabad	Palanpur	1957–93	Bliss and Stern (1982); Lanjouw and Stern (1998)
Uttar Pradesh	Agra	Sunari	1964–85	Fuhs (1988)
Uttar Pradesh	Aligarh	Parhil	1970–87	Saith and Tankha (1992)
Uttar Pradesh	Meerut	Wahidpur, Rampur, Izarpur, Jamalpur	1963–89	Sharma and Poleman (1993)
Uttar Pradesh	Allahabad, Muzaffarnagar	Alipurjeeta, Chhaukra	1964–86, 1977–86	Srivastava (1995, 1997)
West Bengal	Bardhaman, Birbhum, Purulia	Seventy-two unnamed villages	1972–86	Bhattacharya, Chattopadhyay, and Rudra (1987a, 1987b)
West Bengal	Bardhaman	Seven unnamed villages	1960s–1980s	Chattopadhyay (1992)
West Bengal	Midnapore	Fonogram, Bidhigram, Keshipur	1986–88	Beck (1994)

[a] In many of these studies, village names have been changed to preserve confidentiality.

Source: Authors' compilation.

We are concerned, in this review, with the evolution of material living standards in rural India. We trace not only changes in poverty or wealth but also the forces that may have governed these changes, such as agricultural intensification, changes in land relations, and occupational diversification. The first section gives a brief account of agricultural intensification and then examines land and labor—two key determinants of rural income—and the evolution of their links to living standards. The second section looks at changes in levels of poverty. Village studies support the common view that poverty has been declining in rural areas but also cite cases of clear impoverishment over time. Villagers' own perceptions of poverty are discussed in this section. The third section examines changes in inequality, the nature and degree of which are difficult to gauge from large-scale statistical data. Village studies suggest that there has been little or no decline in the economic stratification of Indian villages. The fourth section summarizes the conclusions from our review and discusses policy implications. Because relatively few village studies set out explicitly to evaluate policy, they offer little guidance for the design of specific interventions. An important lesson that emerges from our review, however, is that inequality at the village level can significantly affect the way policies operate in practice.

The Sources of Income

In most of India's half million or so villages, agriculture remains central to the local economy. The intensification of agriculture that began with the green revolution during the 1960s—the introduction of new fertilizers and seeds and expanded application of existing inputs, such as irrigation and labor—has had a profound impact in many parts of the country (Griffin 1979; Lipton and Longhurst 1989; Singh 1990; Hazell and Ramasamy 1991). The degree to which agricultural practices have changed has varied markedly by region and agroclimatic zone, but few rural areas have remained unaffected.

During the early years of the green revolution, there was much debate about its distributional impact. Many observers thought it could lead to rising income inequality. They reasoned that large farmers, with lower risk aversion and better access to credit, would make more extensive use of the new technologies and that the greater economies of scale in the new farming methods would enable them to increase their agricultural income far more than small farmers could, thus widening the income gap between the two groups. It was further suggested that inequality in land ownership, and even landlessness, would increase as the early beneficiaries of the green revolution, the large farmers, bought land from smaller farmers. We consider below the extent to which village studies support these concerns.

Agricultural Intensification

Virtually all of the studies surveyed find some expansion of agricultural production since the late 1950s. In Palanpur, Uttar Pradesh, productivity improved significantly between 1957–58 and 1993, and particularly after the mid-1960s. Output per acre of wheat, the main food crop in Palanpur, increased by about four times during the period, an annualized per capita growth rate of about 2 percent (Bliss, Lanjouw, and Stern 1998; see also Saith and Tankha 1992, for another village in Uttar Pradesh, and Leaf 1983, for Punjab). Although the most significant growth in output is seen in the northern states of Haryana, Punjab, and western Uttar Pradesh, substantial growth occurred elsewhere as well (for Tamil Nadu, see Gough 1981; Guhan and Mencher 1983; Athreya, Djurfeldt, and Lindberg 1990; Ramachandran 1990; Harriss 1991; and Hazell and Ramasamy 1991). Even some of the semi-arid regions in the ICRISAT studies —Mahbubnagar, in Andhra Pradesh, and Sholapur and Akola, in Maharashtra— showed increasing agricultural intensification between 1975–76 and 1983–84 (see Walker and Ryan 1990, for a review of these remarkable studies sponsored by the International Crop Research Institute for the Semi-Arid Tropics).

Irrigation, which permits multiple cropping and a shift to high-value crops, appears to have made the greatest contribution to agricultural intensification. Saith and Tankha (1992) note that in Parhil, Uttar Pradesh, farmers began harvesting three crops a year by 1987 and that high-value crops such as vegetables had replaced mixed crops. In Palanpur, where virtually all cultivated land was irrigated by the early 1980s, mixed crops and coarse cereals yielded to wheat, rice, and sugarcane (Bliss, Lanjouw, and Stern 1998). Similarly, in Iruvelpattu, Tamil Nadu, villagers were harvesting three crops a year by the early 1980s, including two paddy crops (Guhan and Mencher 1983).

Although relatively few studies explicitly compare output per hectare and farm size, those that do find little evidence of economies of scale. In Palanpur, output per hectare was found to be unrelated to farm size in any of the survey years (Bliss, Lanjouw, and Stern 1998), and in North Arcot district, Tamil Nadu, the new technologies appear to have been broadly "scale-neutral" (Hazell and Ramasamy 1991).

Land Ownership and Tenancy Relations

INCIDENCE OF LANDLESSNESS. Statistical evidence on trends in landlessness in rural India is scant and sometimes inconsistent (see Raj 1976). The number of village studies showing a drop in landlessness, however, is surprising in light of the popular presumption that agricultural intensification has caused it to rise.

Moreover, village studies suggest that impoverishment is only one of the causes of landlessness. Others include changes in household structure, population growth,

migration, and occupational change. Thus, increased landlessness, even where it i observed, need not be associated with worsening economic conditions.

Rodgers (1983) reports net losses in land owned and sharecropped in 46 smal and marginal farming households in northeast Bihar between 1971 and 1981. Ye landholdings and local wages were not the sole determinants of household incomes which actually rose slightly over the period. This was in part attributable to the growth in remittances from family members in Assam. In such cases, a "decline" into landlessness might indicate that households can afford to withdraw from agriculture altogether and might therefore imply an improvement in living standards.

Attwood (1979), using retrospective data for a village in Maharashtra's famine tract finds that the proportion of landless households increased between 1920 and 1970 but that this increase was caused mainly by in-migration. The availability of nonfarm employment in the local cooperative sugar factory made it possible for these landless immigrants to enjoy a living standard comparable to that of the landed population.

Drèze, Lanjouw, and Sharma (1998) show that landlessness in Palanpur was about 14 percent between 1957–58 and 1974–75, rose to 23 percent by 1993, and fluctuated considerably over time across households. Very few households became landless because the land was sold, however. More commonly, this occurred when sons left their father's household before his death, prior to inheriting their share of the land. Much of the movement *out* of landlessness in Palanpur then occurred when these sons subsequently acquired their inheritance. In Gokilapuram, Tamil Nadu, Swaminathan (1991) notes high immobility in land ownership at either end of the landholding scale. Once a household is landless, it tends to remain so, but in between the extremes, and particularly among holders of 2.5 to 5 acres of land, there is considerable upward and downward mobility.

In the ICRISAT villages described by Walker and Ryan (1990), the general trend from 1950 to 1982 has been toward decreased landlessness in areas of reliable rainfall, increased equality of landholdings, and a decline in average farm size (see also Gadre, Wahile, and Galgalikar 1987). Athreya, Djurfeldt, and Lindberg (1990) note that in Thiruchirapalli district, Tamil Nadu, smallholder cultivation increased, landlessness decreased, and inequality in landholdings declined over the last generation. In irrigated areas, the incidence of landlessness decreased from 64 to 55 percent, and the land area occupied by landholdings of 25 or more acres decreased from one-half to one-third of the total land area. In the dry areas, half of those who were landless acquired some land. In Karimpur, Uttar Pradesh, the proportion of landless families decreased from 1925 to 1975, with a further sharp drop between 1975 and 1984 (Wadley and Derr 1989). In North Arcot, Tamil Nadu, Hazell and Ramasamy (1991) find no evidence of increased loss of land by smallholders over time.

LAND TRANSACTIONS. Srinivas (1976:82) reports that in Mysore, "parting with ancestral land was a serious matter under any circumstances." The low levels of activity

in land markets in most of the villages studied suggest that this attitude may well generalize to rural India more broadly (and to other developing countries as well: see Binswanger and Rosenzweig 1986a).

In Palanpur, Drèze, Lanjouw, and Sharma (1998) find that the land market is quite inactive in the sense that sales and purchases of land occur rarely. From 1957 to 1993, the average amount of land sold in Palanpur was barely 0.5 percent a year. Analogous results are reported for Parhil (0.5 percent) by Saith and Tankha (1992), for the ICRISAT villages (0.7 percent) by Walker and Ryan (1990), and for Konogram (less than 1 percent) by Beck (1994). Bliss, Lanjouw, and Stern (1998) suggest that typical motives for land transactions in industrial countries, such as life-cycle changes, moving to a job, and changes in perceived returns to different forms of assets, are muted in rural Indian villages. In addition, the "thinness" of the land market in villages such as Palanpur can be linked to a number of different kinds of transaction costs.

Although *inactivity* in the land market challenges the notion of rapid land polarization in rural areas, *activity* need not imply increased polarization. Harriss (1991) finds that small landowners in North Arcot district frequently added to their holdings, whereas large landowners generally sold land. In Palanpur, most of the land sold between 1957 and 1993 was sold by households of the Thakur caste, which was the dominant landowning caste in the village, but one whose traditional occupation was not cultivation (Drèze, Lanjouw, and Sharma 1998). A similar pattern of sales has been noted by Jha (1994) in Bihar.

Some active land markets do appear. Epstein (1973) reports that the conversion of dry land to wet (through canal irrigation) in rural Karnataka led to an "overnight" increase in land prices, presumably because the sudden arrival of irrigation led to a rapid change in land productivity and the expected returns to land ownership. Early land sales proved to be economically disadvantageous for the sellers, and those who had better access to credit and were less risk-averse were able to buy this land and profit in the long term.

Irrigation is often accompanied by a shift to cash crops which, although potentially lucrative and possibly less exposed to harvest fluctuations than nonirrigated crops are vulnerable to the vagaries of the market. Booms and busts in the sugar market were a driving force behind distress sales in the Maharashtra village studied by Attwood (1979) and were an important cause of downward mobility among the largest cane growers and their creditors. In this particular case, the sales had an equalizing effect on the distribution of land ownership.

LAND LEGISLATION. Village studies suggest that land legislation directed toward limiting the size of landholdings or securing sharecroppers' tenure has had mixed results. Most studies find that the ultimate changes induced by land legislation are not those intended by its architects.

Athreya, Djurfeldt, and Lindberg (1990) conclude that in Thiruchirapalli, land-reform legislation compelling large landlords to sell holdings resulted in a more equal distribution of land. They attribute this result to the particularly active tenants movement in the area, combined with a high incidence of absentee landlords. The experience elsewhere in Tamil Nadu was not so positive, however. In Tanjavur district, Gough (1987) finds that land reform led to the widespread evasion of land ceilings and to the eviction of tenants. Kapadia (1993) reports a similar pattern among the Pallars in Poovaloor.

In neighboring Kerala, Mencher (1980) finds that the 1970s land legislation giving permanent tenure and ownership rights to former tenants had two effects. Landlords who had failed to rid themselves of tenants prior to the legislation lost land because of it, but the only benefit most agricultural laborers gained was the permanent right to their house sites and to a tiny piece of the immediately surrounding land.

In the semi-arid ICRISAT villages, little land changed hands directly because of land-ceiling and tenancy legislation, but because the threat of confiscation was perceived as real, large farmers saw an increased risk in land accumulation (Walker and Ryan 1990). This perceived risk, coupled with the advent of irrigation, caused a shift away from the acquisition of rainfed land to investment in irrigation. Rather than evicting tenants, large farmers in the ICRISAT villages seem to have developed an aversion to long-term tenancy—a development that Walker and Ryan claim has led to more equitable access to land among prospective tenants.

In Karimpur, Uttar Pradesh, two developments changed the pattern of land distribution (Wadley and Derr 1989). The first, the Uttar Pradesh Zamindari Abolition and Land Reforms Act of 1950, combined with land ceilings to eliminate absentee land agents (zamindars) and give land ownership to the tillers and also to redistribute land from Brahmans to non-Brahmans. The results were by and large progressive. The subsequent land consolidation during 1967–70, however, although meant to regroup the holdings of small farmers, seems to have benefited richer farmers disproportionately. Drèze, Lanjouw, and Sharma (1998) note that in Palanpur an attempt at very modest land redistribution in 1976 (six households were allotted one acre each of government-owned Palanpur land) involved considerable corruption and did not benefit the poor. A land-consolidation exercise carried out in Palanpur in 1985–86 was broadly successful in reducing land fragmentation, but influential farmers exploited their connections with the headman to get the best land.

In West Bengal, Chattopadhyay (1992) claims that land-ceiling laws had no direct impact on the village of Raj●or because no land was declared as surplus and vested with the government for redistribution. He finds, however, two important indirect effects. First, large joint families, in an attempt to evade the land ceilings separated into smaller proprietary units, an action that led to fragmentation of large families and the erosion of their dominance in village politics. Second, a

observed elsewhere, the land-ceiling legislation led to the large-scale eviction of tenants.

TENANCY ARRANGEMENTS. Village studies provide ample evidence of the resilience of tenancy as an institution. They point to three broad developments over time: a move from sharecropping to fixed-rent contracts, including agreements for payment in kind (which might be misconstrued as sharecropping); a greater involvement by landlords in cost-sharing and in the supervision of day-to-day operations; and an emergence of "reverse tenancy," that is, the leasing of land by larger landholders from smaller landholders.

In Palanpur, the tenancy market, which is very active, has several notable features. First, with the exception of the poor, who may be more excluded than in the past, both tenants and landlords appear remarkably similar on average and are widely distributed along the scales of per capita income, land ownership, and caste status. Second, there is considerable turnover each year, with landlords and tenants re-sorting themselves and sometimes changing roles (Drèze and Sharma, 1996).

The evolution of tenancy contracts in Palanpur has been closely associated with the importance of nonlabor inputs (irrigation, fertilizers, seeds) accompanying the intensification and increased market orientation of agriculture. Cost-sharing of inputs, which has become more common in sharecropping contracts, has led to a reduction in economic differentiation between landlords and tenants, as poor households lose the ability to contribute their share of the cultivation costs. A similar trend is noted in Tamil Nadu by Ramachandran (1990) and in Uttar Pradesh by Srivastava (1995). Furthermore, ownership of indivisible productive assets such as pumpsets or tractors induces some of the larger landowners to lease-in land, giving rise to the phenomenon of reverse tenancy (Walker and Ryan, 1990; Janakarajan, 1996). The emergence of tenants who are less risk-averse and credit-constrained than in the past also explains the shift away from sharecropping contracts to fixed-rent tenancy contracts in Palanpur.

Occupational Change

DECLINE OF TRADITIONAL LABOR SERVICES. Many village studies observe a decline in traditional caste occupations. Wadley and Derr (1989) note that the Hindu jajmani system, in which customary payments are received in return for regular service to a patron, became virtually extinct in Karimpur, Uttar Pradesh, between 1925 and 1984. They attribute the loss of demand for these services to three main factors: behavioral change, technological change and mechanization, and increased monetization of transactions.

Athreya, Djurfeldt, and Lindberg (1990) comment on the extinction of some services and increased casualization of others in Tamil Nadu. They report that even

those traditional artisans who continue to be regulated by the old jajmani syst often earn cash from some customers and work as agricultural wage laborers. L (1983) finds that, in Ludhiana, Punjab, most harijans, who were formerly weave are now full-time agricultural workers. Some traditional occupations, however, su as carpentry, continue to be in strong demand (Drèze, Lanjouw, and Sharma 199

Ramachandran (1990) points out that barriers to employment in agriculture a negligible. In Gokilapuram, Tamil Nadu, two-thirds of the service-caste memb have moved out of their traditional activities into agriculture. Indeed, agricultu labor has become the most caste-heterogeneous activity in the village.

THE CASUALIZATION OF LABOR CONTRACTS. Village studies also note a decline traditional farm-labor arrangements. Long-term relationships between employers a laborers have declined in favor of casual, nonpersonalized, contracts, and the su contracting of specific cultivation tasks to labor "gangs" has increased.

Decline is not the same as demise, however. In some cases, the traditional fa servant arrangement has been replaced by the "right of first call," whereby work first check at their patron's house to see if their services are needed before seeki employment elsewhere. For Ramachandran (1990), the right of first call remains manifestation of "unfreedom" for agricultural laborers. Epstein (1973), howev recounts that casual farm workers in unirrigated Dalena expressed envy at the co tinuation of the permanent farm-servant institution, however "unfree," in neighbo ing Wangala.

In Thanjavur, Tamil Nadu, Gough (1987) observes a movement over time aw from in-kind payments. In the 1950s, attached laborers were generally given clot ing, life-cycle-rites goods, and a plot of land; in the 1980s, they were given no la and received more of their income in cash. Walker and Ryan (1990) find a simil decline of in-kind payments to farm servants in several of the ICRISAT villages. On in Sholapur, which is drought-prone and characterized by high income variability payment in kind still important.

The increase in casualized labor contracts and cash payments suggests that wage may reflect valuable distributional information. Because agricultural daily-wag rates are remarkably uniform for all laborers (of the same gender) within a villag wage trends should show whether laborers' incomes are rising over time— assumin no offsetting changes in days of employment (Drèze and Mukherjee 1989, Da 1996).

Two additional trends are observed in village studies. First, permanent servant are being replaced by gang (often migrant) labor, and second, piece-rate contract have increased over time. Indeed, gang labor, in which job payment is shared among members of the gang, is a collective piece-rate system. Breman (1993) claims that th movement after World War II from cane cultivation (which is highly labor inten sive) to fruit tree cultivation (which requires hired labor only during the picking

season) was the primary cause for the expansion of gang labor in Gujarat. Athreya, Djurfeldt, and Lindberg (1990) find that in Thiruchirapalli, Tamil Nadu, contract gang labor has been gradually replacing wage labor, even for labor-intensive crops. Leaf (1983) notes that in Punjab, the green revolution has led to a marked rise in "periodic labor needs," which has been met by an increased use of gang labor (see also Desai 1983, and Saith and Tankha 1992).

Drèze and Sharma (1996) note that in Palanpur, farmers have a clear view of the respective merits of piece-rate and daily-wage contracts. Piece-rate contracts dispense with the need for close supervision and encourage a timely completion of tasks. They are well suited to activities such as harvesting but may be less appropriate for tasks requiring quality control. They may also be appealing to highly productive workers, who can raise their earnings above the daily-wage rate through piece-rate work, as well as to the least productive workers, who may be excluded from the daily-wage labor market (Baland, Drèze, and Leruth 1996).

EXPANSION OF NONAGRICULTURAL EMPLOYMENT. Discussions of occupational change in village India tend to concentrate on the decline of traditional labor services. This trend is often viewed negatively, on the assumption that workers are being "pushed" out of traditional occupations. A relevant but little discussed topic is the expansion and diversification of nonagricultural employment.

Many village studies note an increase in nonagricultural employment. Wiser and Wiser (1971) mention a bus stand tea stall and new bicycle and tractor repair shops. Epstein (1973) reports on the movement of entrepreneurs in 1970 to cafés, shops, cattle-trading posts, cane crushers, and rice mills, none of which had existed in 1955. Srinivas (1976) notes investment in bus lines, and Saith and Tankha (1992) comment on band-playing as a specialty of growing importance in Pathil, Uttar Pradesh. Although these are all instances of self-employment, evidence suggests that nonagricultural wage employment has expanded even more rapidly than nonagricultural self-employment (Visaria and Basant 1994). Because these new employment opportunities tend to be caste heterogeneous, they compensate in part for the reduced market for traditional caste-specific labor services.

Nonagricultural labor-market opportunities appear to be an important means of offsetting declines or high variances in village incomes. In North Arcot, Tamil Nadu, new agricultural machinery displaced hired labor in paddy cultivation, but real-wage rates in agriculture rose (at least for some activities) as a result of increased nonagricultural employment and the consequent tightening of village labor markets (Hazell and Ramasamy 1991). In the ICRISAT villages, nonagricultural earnings became increasingly important sources of income in the 1980s, raising mean household income and dampening variability (Walker and Ryan 1990). Decreased variability, in particular, seems important in raising rural living standards.

In Palanpur, both regular and casual employment outside the village has expan (Bliss, Lanjouw, and Stern 1998; Drèze, Lanjouw, and Sharma 1998). The dem for employment in this sector, where wage rates and work conditions are attrac relative to agricultural work, exceeds the supply of jobs available. The process thro which these jobs are obtained appears to be governed by both personal connect and the ability to pay a bribe. Regular nonagricultural jobs tend to cluster in a sr number of establishments where an initial employee has helped others to gain ployment. Those who follow the first entrant are frequently of the same caste or otherwise related.

The role of personal contacts and influence in job search might explain the l gaps between agricultural and regular nonagricultural wages, the low turnove regular nonagricultural jobs, and the apparent disadvantage that persons of low cial status have in competing for regular nonagricultural jobs, even given compa skills and endowments (Unni 1997). Although the better educated, or other privileged, may have more opportunities for nonagricultural employment, the p with lower reservation wages, have in the past been most active in pursuing non cultural opportunities. This appears to be changing, however; the village elite now aggressively seeking nonagricultural employment. In Palanpur, there has be clear shift over time in the distribution of nonagricultural earnings, with the bet off in the village acquiring an increasing share. This gradual reduction in the sh for disadvantaged groups has also been observed by Leaf (1983) in the Punjab and Wadley and Derr (1989) in western Uttar Pradesh (for which, see also Sharma Poleman 1993 and Ranjan 1994).

AGRICULTURAL WAGES. The green revolution and the expansion of the nonf economy appear to have raised agricultural wages in rural India by increasing demand for labor. Until recently, secondary data suggested that real wages in r India showed no significant upward trend (Kurien 1980). Evidence now sugge however, that wages did rise in most regions of India in the 1970s and 1980s (Ach 1989; also Guhan and Mencher 1983; Leaf 1983; Harriss 1989; Ramachandr 1990)—although there is no clear trend in the 1990s (Unni 1996).

In Palanpur, real wages have risen fairly steadily since 1974–75, remaining w above the levels that prevailed in either 1957–58 or 1962–63 (Drèze, Lanjouw, a Sharma 1998). This increase is all the more dramatic when expressed in term wheat purchasing power. Because the relative price of wheat fell in Palanpur as pr duction rose over time, one day of casual labor could purchase more than 8 ki grams of wheat in 1993, compared with less than 3 kilograms in 1957–58. Simila Hazell and Ramasamy (1991) show that, in North Arcot, agricultural earnings f landless laborers, small paddy farmers, and nonagricultural households doubled l tween 1974–75 and 1983–84, as large farmers withdrew from agricultural labor a employment opportunities in dairying and nonfarm activities expanded.

OR MIGRATION. The rise in work-related migration also affects village incomes. ig large-scale survey analysis, Kurien (1980) finds that in rural Tamil Nadu, the ively poor seem increasingly willing to give up not only their traditional occupa-s, but also their places of work. Walker and Ryan (1990) observe a tightening of abor market in Aurepalle, Andhra Pradesh, from 1974 to 1985 and a 60 percent in wage rates, which they attribute to temporary migration to nonagricultural in Hyderabad. Saith and Tankha (1992) find that in Parhil, Uttar Pradesh, the dence of out-migration virtually doubled.

he effects of labor migration are difficult to examine in a closed village study see Breman 1985; Walker and Ryan 1990; and Hazell and Ramasamy 1991). ertheless, village studies do note an apparent shift from low-caste, low-skilled ation to high-caste, high-skilled migration. In Karimpur, extravillage service had formerly been held solely by the outcaste poor, who migrated to Calcutta to k as sweepers. By 1984, service jobs had become the "desired occupation" of thier Jati families (Wadley and Derr 1989). In Punjab, as well, high-caste Jats migrated out more frequently since the green revolution than have outcaste jans (Leaf 1983).

LETARIANIZATION? Evidence from secondary sources, such as the census, sug-s that an increasing proportion of the rural population is working for wages. This ease has sometimes been seen as a trend toward "proletarianization" of the labor e. This term can be interpreted in (at least) two ways. It may simply describe the away from self-employment (mainly in agriculture) to wage labor, or it may gest that smallholder cultivators are being pushed out of agriculture into wage r, rather than being pulled by new employment opportunities.

Harriss (1991) argues that in North Arcot, Tamil Nadu, proletarianization is mainly ocess of farming households supplementing their cultivation incomes with wages n both agricultural and nonagricultural employment. He interprets a sharp in-se in the number of occupations outside of cultivation or agricultural labor be-en 1973 and 1984 as a process of "proletarianization without depeasantization" riss, 1989; but compare Athreya, Djurfeldt, and Lindberg 1990, who describe ocess of "peasantization rather than proletarianization" in Thiruchirapalli). achandran (1990) also demonstrates that a rise in agricultural wage employ-it does not necessarily imply a withdrawal from cultivation. He finds that in alapuram, Tamil Nadu, smallholders supplemented their cultivation activities agricultural employment but did not sell their own land. He therefore suggests the expansion of the market for hired labor has helped to preserve the institu-of smallholder cultivation.

he degree to which expanded agricultural wage employment is associated with a idrawal from cultivation seems to be closely linked both to the incidence of ten-evictions accompanying land-to-the-tiller legislation in the 1970s and to dimin-

ishing farm size caused by population pressure on the land. In Thanjavur, Ta
Nadu, the number of landless male agricultural laborers pushed out of cultivat
for these reasons increased dramatically (Gough 1987). Workers have also been pu
away from farms by increased nonagricultural opportunities. Whether push or p
influences predominate across rural India varies with the particular local experie

SOCIAL IMPLICATIONS OF OCCUPATIONAL CHANGE. Occupational change is likely
be accompanied by changes in the prevailing social order. Breman (1993:21),
instance, regards the disintegration of the jajmani system as a product of multi
factors and states that "the relationships lost their local flavor in the process of
largement of scale. Commercialization of agriculture and continuously increas
government intervention diminished the importance of local autarky and autonor
The drift of members of the artisan and serving castes to the urban centres, fr
which they began to serve the surrounding countryside, contributed to the rise of
ever-increasing number of different contacts which went beyond the village. I
pending on the accessibility of the region, this development began early or late

The degree to which a village is linked to the market economy can significar
affect its occupational and social structure. Gough (1981), in her analysis of soci
change in Thanjavur, Tamil Nadu, remarks that coastal regions, which have be
more "disturbed" than their inland counterparts by external change, break mo
easily with traditional caste-based occupational structures. Where an area's compa
tive advantage lies in agriculture, however, the traditional social order (and asso
ated occupational structure) is less likely to change. Epstein (1973) describes a si
lar process in Karnataka. One of the villages studied, Wangala, was served by a rece
constructed irrigation canal; the other, Dalena, was not. Irrigation in Wangala e
abled farmers to grow cash crops, but because the village's economy remained alm
wholly agricultural, the new cash economy coexisted with the traditional system
hereditary relationships between Wangala farmers and their "functionaries." I
system was, in fact, reinforced by the introduction of labor-intensive cash crops
neighboring, unirrigated, Dalena, economic diversification led to increased facti
alism. Agriculture no longer bound the villagers together, and greater integra
into the rural economy led Dalena commuters into the wider world, where differ
tiation along caste lines and other social institutions were diluted.

Poverty and Living Standards

General Trends

Although aggregate economic statistics in India point to a steady (albeit slow) de
cline in poverty, absolute levels of deprivation remain high. Few village studies tra

absolute poverty, measured in terms of consumption or income. They generally scrutinize, instead, the positions of households relative to one another. Beck (1994) thus argues that poverty rose in three West Bengal villages during the late 1980s, even though the incomes of the poorest households increased. His judgment, reflecting the perceptions of villagers themselves, is based on the observation that the rich in these villages enjoyed significantly larger increases in income than the poor did during the same period.

Village studies that do focus on absolute poverty also observe a decline. In Palanpur, the proportion of the population below any reasonable poverty line clearly decreased between 1957–58 and 1983–84, although poverty rates fluctuated markedly throughout the period, and households moved in and out of poverty in response to price changes, harvest levels, and the partitioning of household lands (Drèze, Lanjouw, and Sharma 1998). In any given year, however, irrespective of the overall level of poverty, households of the low-ranked Jatab caste, and households relying primarily on casual agricultural work for income, were disproportionately represented among the poor. This relatively high and constant risk of poverty among agricultural laborers and low-caste households is stressed in many village studies (Mencher 1980; Rodgers 1983; Ramachandran 1990; Jha 1994). Drèze, Lanjouw, and Sharma (1998) argue that because agricultural wage labor is seen as a last resort by Palanpur villagers, it is a reliable indicator of poverty in any year.

Walker and Ryan (1990) find that two-thirds of the villagers in the ICRISAT studies moved in or out of poverty in at least one of the nine consecutive years between 1975–76 and 1983–84. These were nearly all middle-size cultivators. The nonpoor—those who never crossed the poverty line—were more educated, did not actively participate in the casual labor market, and owned more land than their neighbors. The consistently poor were predominantly landless harijans with high dependency ratios.

Village studies recording income levels for more than two years are rare. The more usual comparisons of income across two periods, corresponding to an initial study and a revisit, are vulnerable to the effects of harvest fluctuations. Where comparisons of income are handicapped in this way, however, other indicators can be scrutinized to assess changes in longer-term living standards. Jodha (1989), using the more "conventional" per capita income measure, finds that 20 percent of the households in his two Rajasthani villages were poorer in 1982–84 than in 1964–65 (although he does recognize that crop yields were better on average in 1964–65). The proportion of households below the poverty line increased from [illegible] to 23 percent, with both upward and downward mobility across the poverty line. Households that appear to have become poorer according to the income measure did seem to be better off, however, when judged by other, qualitative, indicators of economic well-being. Jodha points to expanded alternatives for employment and borrowing, to an increase in consumption activities with high income

elasticities (travel, slack-season purchases), and to investment in lumpy consum durables (pukka houses and structures).

Increased ownership of consumer durables has been observed in many villa studies. Leaf (1983), for example, finds that in rural Punjab, a substantial improv ment in housing and shelter accompanied the green revolution. Bhattachary Chattopadhyay, and Rudra (1987a) note that in rural West Bengal, ownership nontraditional items such as radios, wristwatches, and bicycles increased conside ably between 1972–74 and 1985–86. Although such increases have been wide interpreted to imply an improvement in living standards, it is important to note th expanded ownership of "modern" consumer durables (the increase most often cite is in part a relative-price effect. In Palanpur, for example, modern consumer durabl have become cheaper relative to livestock and have increasingly been substituted fo livestock in dowries (Drèze, Lanjouw, and Sharma 1998). Similarly, an increase consumption of goat meat in Punjab can be attributed both to an improvement general welfare and to the pumpset revolution, which led to a shift from own draft animals to maintaining livestock for food and marketing purposes (Leaf 1983

Instances of Impoverishment

Although both large-scale surveys and village studies find a general drop in pover after the 1960s, there are pockets where poverty has increased. These instances impoverishment are worth scrutinizing.

Rapid population growth, which can offset rising productivity and reduce pe capita income, is one factor that can cause impoverishment. Village studies note th growth in the agricultural labor force may follow land-augmenting technologic change, particularly when such change induces in-migration of landless laborin households. Ramachandran (1990) observes a decline in wage rates in Gokilapura between 1948 and 1975, which he attributes to a sharp growth in the relative an absolute size of the agricultural labor force in the Cumbum valley. Walker and Rya (1990) find that in Kanzara, the influx of landless laboring households from neig boring areas has kept wages low, despite the increased demand for agricultural labo Where there are factors pushing smallholders into landlessness and agricultural la bor, the need for offsetting "pull" factors becomes paramount. As Jha (1994) ob serves, "push" sometimes dominates. Population growth, tenant eviction, and de clining migration opportunities in Bihar have led to a decrease in employment day per worker and to stagnating incomes for daily-wage and attached laborers.

Many parts of rural India, moreover, are vulnerable to drought, and drought ca devastate the poor. Hazell and Ramasamy (1991) show that in North Arcot, Tami Nadu, average incomes in 1982–83 (a severe drought year) were lower than in 1973– 74 for all agriculturally dependent households. Many households in North Arcot had no financial reserves with which to cope with droughts, particularly when low

rainfall years ran together. The expansion of irrigation did not exercise a stabilizing effect in North Arcot because irrigation water comes from groundwater reserves, which depend on rainfall for their recharge. Although groundwater irrigation is a useful way of redistributing the monsoon rains in this region, it provides only limited capacity for carrying water from good to bad years. The close association between poverty and access to groundwater is explored by Bhatia (1992) for rural Gujarat and by Saith and Tankha (1992) and Drèze, Lanjouw, and Sharma (1998), who express concern that unregulated and expanding pumpset irrigation is depleting groundwater in the Gangetic plain.

Even against a background of generally declining poverty, certain subgroups of the population face a high risk of impoverishment through illness, accidents, or life-cycle events. Drèze (1990) notes, for example, that in most parts of India, women, who are usually younger than their husbands, face a high probability of becoming widowed during their lifetime. Whether widowhood translates into a sharp decline in their living standards depends on local inheritance laws, whether women are allowed to work for wages, and whether the widow has children (in particular, sons) who can provide support.

The Perceptions of the Poor

Income- or consumption-based measures of well-being are often suspected of failing to capture local perceptions about living standards. Village studies are particularly useful in this regard because the authors are close to their subjects. Bhattacharya, Chattopadhyay, and Rudra (1987b) find that 60 to 80 percent of the households they canvassed in rural West Bengal felt that their standard of living had improved with respect to *social consumption* (drinking water, medical care, education, roads and transport, and recreation and culture), and less than 10 percent felt it had deteriorated. More generally, Bhattacharya, Chattopadhyay, and Rudra (1987b) find that about half of the respondents felt their overall living standards had improved, while about a quarter thought they had deteriorated. The main reason for their judgments was a perceived increase or decrease in real income.

In his study in western Rajasthan, Jodha (1989) points to reduced reliance on traditional patrons as a reason for a perceived improvement in well-being. The importance attached to personal freedom and dignity is a recurrent theme, even among the very poor, and a desire to secure such freedom can be a driving force behind the acquisition of assets. Srinivas (1976:111) writes that "landownership and wealth were occasionally able to mitigate if not overcome the effects of birth in a ritually low caste. . . . No wonder then, that there was a general scramble for land . . . for it meant freedom from hunger and bondage to patron, and also self-respect." Wadley and Derr (1989) find that in Uttar Pradesh, those who perceived an improvement in their conditions spoke of it in terms of increased personal freedom rather than re-

duced hunger or poverty. Beck (1994) reports that in West Bengal, 49 out of respondents claimed they valued self-respect more than food.

Some think otherwise, however. Epstein (1973) quotes a Dalena villager as s ing, "you cannot eat social acceptance." Breman (1993) claims that in Gujarat, b landowners and agricultural laborers preferred long-term labor contracts. Such co tracts guaranteed employment for laborers during the slack season and laborers landowners during the peak season. That concerns about personal freedom can something of a luxury is starkly pointed out by Mencher (1980), who warns agai complacency regarding rural Kerala, where the remarkable inroads achieved in hea and literacy might suggest that poverty has been eliminated. Mencher argues that t "miracle of Kerala" has not been a miracle for agricultural laborers, many of th women, who are still daily confronted with the uncertainty of how they are going feed their children.

Inequality

Changes in inequality are closely related to the manner in which rural poverty l evolved. Understanding the distribution of living standards in village communit can therefore provide important insights for poverty-reduction initiatives. There ha been many efforts to strengthen local decisionmaking power in rural India, fueled the impression that local control will contribute to poverty reduction. The Jawah Rozgar Yojana (JRY) employment program, for instance, introduced by the cent government in 1989, relies on decentralization. The JRY scheme is implemented l village panchayats and promotes the creation of durable community assets. H much one can expect from such a scheme depends, to some extent, on how w represented the poor are in local decisionmaking. This, and factors such as villa solidarity, are likely to be affected by the degree of polarization in village livin standards.

Because village studies typically take the entire village population as their doma they are better suited than large-scale surveys for studying village-level inequalit Large-scale surveys rarely sample enough households within a given village to yie reliable measures of village inequality. Moreover, because of factors such as aging is risky to make inferences about the evolution of inequality by following a speci sample of households surveyed over time.

Land Ownership

Village study findings on changes in land distribution challenge several comm clichés. First, contrary to conventional wisdom, there seems to be considerable move ment in the distribution of land, movement not necessarily linked to a highly acti

narket. Second, there is no uniform trend toward increased inequality in land ›ution. Whether land inequality has increased or decreased varies across the ›s studied and often depends on whether the area in question is wet or dry.

nographic change is particularly relevant to changes in land distribution. Srinivas :112) writes that "while a man may have had his descendants in mind when g land he also knew that it would be divided after his death. Big estates were usually built up through the accident of single sons in more than one genera- In Palanpur, demographic factors account for the bulk of observed changes in stribution of land (Drèze, Lanjouw, and Sharma 1998). Removing the effect of hold partitioning by aggregating the land ownership of each "dynasty" (de- as the union of all households descending from a particular household in the urvey year, 1957-58) reveals that inequality in per capita land distribution ned fairly constant between 1957–58 and 1993.

minathan (1991), however, finds high and rising inequality in landholdings kilapuram between 1977 and 1985 (despite a decline in the already lower ality in nonland assets). Hazell and Ramasamy (1991) find no evidence of sed inequality in landholdings in North Arcot following population growth rtitioning but note that average farm size declined slightly across all quartiles en 1973 and 1983. In the ICRISAT villages, Walker and Ryan (1990) actually creased equality in land distribution, with large landowners losing ground to landowners in both relative and absolute terms.

: positions of smallholders and the landless appear to be relatively static. Athreya, eldt, and Lindberg (1990) report that in Thiruchirapalli, the most immobile s in their respective areas are the landless in the wet area and those who began ery small holdings in the dry area. Similar patterns are noted by Rao (1972); od (1979); Cain (1981); Caldwell, Reddy, and Caldwell (1982); Harriss (1991); waminathan (1991).

reya, Djurfeldt, and Lindberg (1990) note that the ownership structure is of- xtremely polarized in wet areas but shows relative equality in dry areas. In chirapalli, this pattern has social origins: in the wet area, lower castes were ionally not allowed to own land, whereas in the dry area, land commanded no at all until the 1900s and is still quite cheap. Hazell and Ramasamy (1991) note he only evidence of a mild worsening in land distribution among their Tamil villages was in Duli, which is poorly irrigated. In dry areas, rising inequality not increase poverty because the productive value of the land is already so low.

ne and Wealth

anpur, there is little evidence that the green revolution led to a marked widen- the distribution of income (Lanjouw and Stern 1998a). In fact, income in- lity appears to have declined somewhat between 1957 and 1984. The sharp

expansion of irrigation during the green revolution years to farms other than wealthiest (from 50 percent of the land in 1957–58 to 96 percent by 1974–75) h an important equalizing effect on incomes. Sharma and Poleman (1993) find a sir lar decline in Walidpur, in western Uttar Pradesh, between 1963–64 and 1988–[illegible] Epstein (1973), however, notes that in Karnataka, there was a sharp concentration income between 1957–58 and 1970. The key factor in the villages Epstein stud was access to irrigated land, and this remained highly unequal throughout the inr val. Households with good land endowments became much richer, while agric tural wage earners competed with a sizable pool of migrant workers for low wag

In Palanpur, Drèze, Lanjouw, and Sharma (1998) find both lower income equality across the green revolution period and a nonnegligible widening in the d tribution of wealth between 1962–63 and 1983–84. These findings need not contradictory. Inequality in wealth may rise, even in the presence of an unchang distribution of income, if the savings function is convex at low levels of income th is, the marginal propensity to save rises with income. An equiproportionate incre in income for all households could thus easily lead to some polarization of new weal In Palanpur, the rise in net wealth inequality also reflects a highly uneven accumu tion of liabilities to credit institutions, as well as the particular vulnerability of t poor to corrupt accounting practices within formal credit institutions (Bell a Srinivasan 1985; Bhende 1986; Binswanger and Rosenzweig 1986b; Iqbal 198 Bouman 1989; Krishnan 1990; and Banik 1992).

Swaminathan (1988) also finds, for Gokilapuram between 1977 and 1985 immense disparity in, and perceptible worsening of, the distribution of wealth, d fined as the households' own estimation of the current value of their assets. It worth noting, however, that although the poor in Gokilapuram become relativ worse-off, the per capita wealth of even the poorest households increased in re terms by a factor of about 2 over this eight-year interval.

Whether inequality has increased or decreased over time in rural India is, perhap of secondary importance to the well-documented observation that *levels* of inequa ity still generally remain high. Although various village studies trace alternative pat of income or wealth inequality, all have started from positions of considerable di parity in living standards. Even where inequality seems to have fallen, only a sm fraction of the total has been eliminated over time.

Caste

Just as land and income distribution in rural India has changed in response to popu lation growth, technological change, and occupational diversification, so caste rela tions have changed. Many studies comment on the shift in local perceptions of, an attitudes toward, caste-based social rankings. Drèze, Lanjouw, and Sharma (1998 point out that in Palanpur, people of all castes can now sit together on the sam

ing cots (charpai), and the dominant Thakurs can no longer exercise arbitrary ·ce over the lowest-ranked Jatabs.

The disappearance of many traditional occupations has inevitably undermined : differentiation of castes by behavior and associations. Walker and Ryan (1990) ribe the erosion of the caste hierarchy in the ICRISAT villages to the tendency of ditional service-related castes to supplement their livelihood with agricultural la- r. Epstein (1973) also notes the role of occupational diversification. She writes it in the unirrigated village of Dalena, economic diversification diluted caste lines, ereas in irrigated Wangala, where the agrarian base remained largely intact, the -India antiuntouchability policies were largely ignored.

The presence of upwardly mobile households clearly affects the social hierarchy in illage. In Palanpur, an ongoing rivalry has pitted the Thakurs, the highest-ranked, iditionally noncultivating and nonlaboring caste in the village, against the Muraos, aditionally cultivating caste (Drèze, Lanjouw, and Sharma 1998). The economic itus of the Muraos has risen substantially since the 1950s and has surpassed that of e Thakurs, resulting in a gradually improved social status for the Muraos and an alating political rivalry with the Thakurs.

Caste mobility is most marked at the middle and upper ends of the caste hierar- y. Kapadia (1993) notes that in Poovaloor, Tamil Nadu, the relative economic ower of the Brahmans has declined. Wadley and Derr (1989) find a similar decline n Karimpur, and Fuhs (1988) observes that in Sunari, Uttar Pradesh, many Brah- mans now even work as laborers for Jat farmers. Breman (1993) finds considerable upward mobility among the middle caste Kolis and Dhodias in two villages in Gujarat. ar (1994) finds similar mobility in rural Bihar, as does Da Corta (1993) in Andhra Pradesh.

Upward mobility of low-caste households is more unusual. Drèze, Lanjouw, and Sharma (1998) find no change in the position of Jatabs in Palanpur society, where they represent about 12 percent of the village population. Jatabs have experienced lower per capita income growth than other groups, almost stagnant levels of educa- tion (100 percent female illiteracy in 1993; 88 percent male illiteracy), virtual exclu- sion from regular, nonagricultural, wage employment, and unaltered land endow- ments. This is despite their being targeted by various government programs. Guhan and Mencher (1983) find that in Iruvelpattu, Tamil Nadu, the combined effects on the harijans of economic inequality, social discrimination, and physical segregation are persistent and readily visible. Epstein (1973) paints a similar picture for Wangala and Dalena in Karnataka.

In some villages, low-caste households have improved their relative standing, through either the influence of (often external) political parties or the forum of col- lective action. Chattopadhyay (1992) attributes the disintegration of the structure of "dominance and subordination" of landless laborers and sharecroppers by three "big- men" families in Bardhaman district, West Bengal, to an awareness of new possibili-

ties fostered by political parties. Gough (1987) attributes the increase in real wage for farm servants in Thanjavur between 1952 and 1976 to the efforts of the communists to ensure good treatment for farm workers, as well as to the rise of labor unions. Sreekumar (1995) cites the importance of collective action in Changel, Bihar, where the consolidation of Yadavs, Noniyas, and Dhanuks, all "backward" castes, led to the break-up of the Khayasth-Brahman monopoly on village politics.

Collective action may not be easily achieved, however. Breman (1993) finds that intercaste tensions among the lower castes in his two Gujarat villages, as well as effective resistance from the higher caste Anavils, have prevented solidarity among the lower castes. Athreya, Djurfeldt, and Lindberg (1990) suggest that in Thiruchirapalli, factionalism among the *higher* castes was a key to reductions in poverty and inequality among the lower castes.

An interesting countertrend to the erosion of caste distinctions is the process called "sanskritization" by Srinivas (1966). Srinivas (p. 28) notes that "when a caste or section of a caste achieved secular power, it usually also tried to acquire the traditional symbols of high status, namely the customs, ritual, ideas, beliefs, and life-style of the locally highest castes." Upwardly mobile castes therefore often come to adhere more, rather than less, stringently to caste norms over time, but they adopt the norms of a caste that is higher than the one from which they originate.

Gender

Although women may have benefited from the general improvement of living standards, their gains have been small compared with the persistent inequalities between men and women in many parts of rural India. In some cases, their relative position has actually declined. This decline is linked in part to the process of sanskritization and the institution of dowry, and in part to the reduced participation of women in the labor force following occupational diversification and technological change.

One of the more alarming findings of village studies has been a decline in the female-male population ratio. In Karimpur, Uttar Pradesh, Wadley and Derr (1989) interpret such a decline among the Jati caste as a growing negative valuation of women linked to the rising incidence in nonagricultural employment outside the village by Jati men. Because female farm workers in Karimpur are generally not hired independently of their husbands, the Jati women now have fewer income-earning opportunities (see also Epstein 1973, for Dalena). Drèze, Lanjouw, and Sharma (1998) link a similar decline in the female-male ratio among Jatabs in Palanpur to the absence of any expansion in female labor-force participation and to a growing identification of disadvantaged castes with the patriarchal norms of privileged castes. Drèze and Sen (1995) record a decline in the female-male ratio among scheduled castes in Uttar Pradesh more generally (from 0.94 in 1901 to 0.88 in 1991).

An important aspect of sanskritization has been the shift from brideprice to dowry. In a study of rural propertied elites northeast of Coimbatore, Tamil Nadu, Heyer (1992) notes a movement from brideprice in the 1930s to dowries in the 1950s and a doubling of the real value of dowries paid from the 1950s to 1970s. Bailey (1957), Epstein (1973), and Wadley and Derr (1989) find similar trends. As Heyer (1992) explains, the growing importance of dowries has led to the redistribution of capital from households with higher daughter-son ratios to households with lower ratios. Maintaining wealth from generation to generation therefore depends on keeping the number of surviving daughters relatively small.

Summary and Policy Implications

Agricultural Production, Land, and Employment

All of the studies reviewed find that agricultural production has grown significantly during the past 20 to 30 years. The broad package of new inputs that accompanied the green revolution has had at least some impact almost everywhere in India. Irrigation, in particular, has had a profound effect.

The new technologies do not appear to have favored large farmers disproportionately. Although large farmers have better access to credit for the purchase of new inputs, the expansion of irrigation may be quite progressive if, prior to expansion, only large farmers had irrigated land.

Landlessness has not obviously increased over time and in some cases has decreased. Because increased destitution is only one possible cause of landlessness, loss of land is not necessarily a symptom of impoverishment.

In most of the villages studied, the land market was found to be relatively inactive. There were exceptions, however. In some instances, canal irrigation, which sharply and dramatically alters returns to land, led to an increase in land transactions. Even in such cases, however, the small and vulnerable farmers were not necessarily the parties losing their land. Land sales, when they occurred, often involved transfers of land from formerly well-endowed, but noncultivating, households to cultivating castes.

Land legislation has had both positive and negative effects on land distribution. The beneficial effects have usually been indirect, such as encouraging farmers to invest in new technologies, rather than to expand their landholdings. A very common response to land-to-the-tiller legislation, however, has been the eviction of tenants. This has increased both landlessness and the size of the agricultural wage-labor force.

Tenancy, where it remains, appears to be evolving along with the new cultivation practices. Many studies point to the frequency of fixed-rent contracts, the greater involvement of landlords in cost-sharing, and the phenomenon of "reverse tenancy."

This evolution of tenancy corresponds to expectations, given the increased intensity of agriculture and the spread of modern practices and technologies. The persistence of tenancy as an institution suggests that it fulfills a useful purpose—permitting the cultivation of land by those who can put it to most productive use, without requiring a full transfer of property rights.

Traditional artisanal occupations have generally declined in number and importance, but this decline has usually occurred against a background of rising real wages. As a result, although most traditional artisans have moved into manual-labor occupations, it is not clear that their living standards have declined. Daily-wage and piece-rate contracts are increasingly replacing long-term agreements. Although such casualization reduces the dependence of laborers on their employers, it can also increase their insecurity.

Rural households benefit from nonagricultural employment not only from the incomes received, but also from reduced exposure to agricultural fluctuations. These benefits have not been shared equally across households, however. Although village households without assets have not been excluded from nonagricultural employment, regular nonfarm employment appears to be linked to personal contacts, as well as to payment of bribes. These factors have probably favored the more affluent and highly ranked villagers, and evidence suggests that this group has become more active in the nonagricultural sector over time.

The expansion of nonagricultural employment opportunities has "tightened" rural agricultural labor markets and raised agricultural wages in many of the villages studied. This increase is less marked in regions where in-migration has risen sharply.

Although proletarianization has been widely observed, it is not clear how many smallholder cultivators are being pushed out of agriculture by declining returns and land polarization, and how many are being pulled away by new opportunities in the nonagricultural sector. The fact that proletarianization often occurs against a background of rising real agricultural wages suggests that pull factors may be just as important as push factors.

In villages having a strong comparative advantage in agriculture, traditional social and occupational structures have generally persisted, even in the face of dramatic changes in terms of trade or technology. In villages that are more heterogeneous economically, however, traditional occupational structures, such as caste-based trades, have become more diffuse, as villagers have acquired links with the outside world and have diversified their activities.

Poverty

Village studies point to a slow decline in rural poverty but note considerable movement in and out of poverty. Some of this movement can be attributed to fluctuations

in harvest quality and to personal calamities. Low caste status and agricultural wage labor are associated with long-term poverty.

Even where poverty does not appear to have fallen, increased ownership of consumer durables, shifts toward consumption of higher quality goods, and self-reported perceptions point to gradually improving living conditions. In addition, the rural poor often single out reduced dependence on patrons as an important improvement in their living conditions.

Although the direction of change is encouraging, poverty remains extremely high in many villages. Perhaps as important to understanding which forces have reduced poverty is the question of why poverty has fallen so slowly in the face of often dramatic growth.

Inequality

In contrast to declining poverty in rural areas, it is difficult to find reduced inequality. Although there has been considerable movement in the distribution of welfare over time, no clear trend emerges, particularly one favoring the very poor.

Land distribution in rural areas is less static than often suggested, but much of the movement may be more apparent than real. Demographic change associated with household partitioning alters the distribution of land but is not likely to reflect fundamental changes in the distribution of wealth. Even where significant changes in land distribution occur, the high degree of immobility among the landless and smallholders suggests that the relatively poor are not the primary beneficiaries. The green revolution may have improved the distribution of welfare, but the scope for improvement depends closely on local conditions.

Rural growth, and in particular the integration of the village economy into the wider rural economy, has eroded the traditional caste structure in many villages. Several studies point to a discernable breakdown in the correlation between ritual status and economic status. In many villages, the traditionally dominant, noncultivating, caste is being overtaken economically by households of other castes, usually traditional cultivators who have taken greatest advantage of the new agricultural opportunities. This process leads, on occasion, to sanskritization, whereby high-caste practices are emulated by other upwardly mobile castes. Although such changes are significant, they do not affect the lives of all castes in the same way. The lowest castes in Indian villages often remain easily distinguishable from the rest of village society by their very low material well-being and the limited opportunities they have to improve their living standards.

In addition, there has been little evidence of improvement in gender relations. The decline in the (already extraordinarily low) female-male population ratio suggests that gender-based inequality may be increasing. Sanskritization may play a

role in this, in that women are often required to observe behaviors associated with higher castes, such as full purdah or withdrawal from wage-labor activity. The shift from brideprice to dowries and the trend toward higher dowries, moreover, further strengthen the perception among households that daughters drain family wealth.

Policy

Conspicuous by its absence from this review is an evaluation of the various public policy measures directed toward village living standards and community life. Although longitudinal village studies seem particularly well-suited to making such an assessment, most of them are silent on this issue.

Drèze, Lanjouw, and Sharma (1998) review every instance of public service provisioning in Palanpur between 1957–58 and 1993—including the building of public schooling facilities, the implementation of the Integrated Rural Development Programme and other credit programs, the JRY employment program, and the provision of widows' pensions. Their conclusion is that, except for the modest success of a program providing two water handpumps near the low-caste quarters, the programs have been extremely disappointing. A recurrent observation is that privileged individuals or groups direct the benefits of programs to themselves at the expense of other village members.

It seems worth asking whether high inequality within villages, and in particular the lack of strong village cohesion, present major obstacles to the successful implementation of policies in Indian villages. The answer could have important implications for the design of policies, especially policies that seek to avoid the recognized pitfalls of centralized delivery by exploiting the potential of decentralized mechanisms. Unless great care is taken to ensure accountability, such decentralized schemes may accomplish no more than the previous efforts.

Notes

Rajshri Jayaraman is a doctoral student in the Department of Economics at Cornell University. Peter Lanjouw is an economist in the Development Economics Research Group of the World Bank and Professor of Economics at the Free University of Amsterdam, Netherlands, where he is currently on leave.

The authors are very grateful to Jean Drèze for guidance in the preparation of this paper, although they do not wish to implicate him for any remaining errors and omissions. They also thank the reviewers for their suggestions, as well as Harold Alderman, Zoubida Allaoua, James Fairhead, Barbara Harriss-White, Jesko Hentschel, Ravi Kanbur, Valerie Kozel, Takashi Kurosawa, Jenny Lanjouw, Michael Lipton, Lant Pritchett, Martin Ravallion, Naresh Sharma, and participants at a seminar held at Queen Elizabeth House, Oxford University, for valuable comments.

eferences

e word "processed" describes informally reproduced works that may not be commonly available ough library systems.

hatya, Sarthi. 1989. "Agricultural Wages in India: A Disaggregated Analysis." *Indian Journal of Agricultural Economics* (April- June):121 -39

hreya, Venkatesh, Goran Djurfeldt, and Stefan Lindberg. 1990. *Barriers Broken: Production Relations and Agrarian Change in Tamil Nadu.* New Delhi: Sage.

twood, D. W 1979 'Why Some of the Poor Get Richer: Economic Change and Mobility in Rural Western India." *Current Anthropology* 20(3):495 -516.

ley, F. G 1957 *Caste and the Economic Frontier* Manchester, U K.: Manchester University Press.

land, Jean-Marie, Jean Drèze, and Luc Leruth 1996. 'Daily Wages and Piece Rates in Agrarian Economies " STICERD Discussion Paper Series DEP 71 London School of Economics, Development Economics Research Programme Processed

nik Arindam. 1992 'Structure of the Credit Market and Its Link with the Other Markets: An Analysis of Village Survey Data in Bangladesh " Delhi School of Economics. Processed.

rdhan Pranab, ed. 1989 *Conversations between Economists and Anthropologists: Methodological Issues in Measuring Economic Change in Rural India* Delhi: Oxford University Press.

ck Tony 1994 *The Experience of Poverty: Fighting for Respect and Resources in Village India.* London Intermediate Technology Publications

ll Clive, and T N. Srinivasan 1985 Transactions in Rural Credit Markets in Bihar and Punjab: An Anatomy ' Working Paper 8 World Bank, Agriculture and Rural Development Department, Washington, D C Processed

ana, Bela 1992 'Lush Fields and Parched Throats. The Political Economy of Groundwater in Gujarat Working Paper 100, United Nations University, World Institute for Development Economics Research (WIDER) Helsinki Processed

attacharya, Nikhilesh, Manabendu Chattopadhyay, and Ashok Rudra. 1987a. 'Changes in Level f Living in Rural West Bengal Housing Conditions ' *Economic and Political Weekly* (September) 1559 66

. 1987b "Changes in Level of Living in Rural West Bengal: Perceptions of the People." *Economic and Political Weekly* (November) 2071 73

inde M J 1986 'Credit Markets in Rural South India ' *Economic and Political Weekly*, Review f Agriculture (September) A119 24

Binswanger, Hans P, and Mark R Rosenzweig 1986a 'Behavioural and Material Determinants of Production Relations in Agriculture ' *Journal of Development Studies* 22(3):503–39.

. 1986b "Credit Markets, Wealth, and Endowments in Rural South India." Paper presented at the 8th World Congress of the International Economic Association, December, New Delhi Processed.

[illegible] J, Peter Lanjouw, and Nicholas Stern 1998. "Population, Outside Employment and Agricultural Change." In Lanjouw and Stern, eds.

[illegible] J, and Nicholas Stern 1982 *Palanpur The Economy of an Indian Village.* Oxford: Oxford University Press.

[illegible]nan, F J A. 1989. *Small, Short, and Unsecured. Informal Rural Finance in India.* Delhi: Oxford University Press.

[illegible]han, Jan. 1985 *Of Peasants, Migrants, and Paupers Rural Labour Circulation and Capitalist Production in West India.* Delhi: Oxford University Press.

———. 1993. *Beyond Patronage and Exploitation: Changing Agrarian Relations in South Gujarat.* Delhi: Oxford University Press.

Cain, Mead. 1981. "Risk and Insurance: Perspectives on Fertility and Agrarian Change in India and Bangladesh." *Population and Development Review* 7(3):435–74.

Caldwell, John C., P. H. Reddy, and Pat Caldwell. 1982. "Perspectives on Fertility and Agrarian Change in India and Bangladesh." *Population and Development Review* 7(3)

Chattopadhyay, S. N. 1992. "Historical Context of Political Change in Rural West Bengal: A Study of Seven Villages in Bardhaman." *Economic and Political Weekly* (March):647–58.

Da Corta, Lucia. 1993. "Household Mobility and Class Differentiation in South Indian Villages." Ph.D. dissertation, University of Oxford. Processed.

Datt, Gaurav. 1996. *Bargaining Power, Wages and Employment. An Analysis of Agricultural Labor Markets in India.* New Delhi: Sage.

Datt, Gaurav, and Martin Ravallion. 1996. "India's Checkered History in the Fight against Po[verty]: Are There Lessons for the Future?" *Economic and Political Weekly* (September) 2479–86

Desai, Vasant. 1983. *A Study of Rural Economies: A Systems Approach.* Bombay: Himalaya Publishing.

Drèze, Jean P. 1990. "Widows in Rural India." STICERD Discussion Paper Series DEP 26. London School of Economics, Development Economics Research Programme. Processed.

Drèze, Jean P., and Anindita Mukherjee. 1989. "Labour Contracts in Rural India: Theories and Evidence." In Sukhamoy Chakravarty, ed., *The Balance between Industry and Agriculture in Economic Development.* Vol. 3. London: Macmillan; New York: St. Martin's.

Drèze, Jean P., and Amartya K. Sen. 1995. *India: Economic Development and Social Opportunity.* Delhi: Oxford University Press.

Drèze, Jean P., and Naresh Sharma. 1996. "Sharecropping in a North Indian Village." *Journal of Development Studies* 33(1):1–39.

———. 1998. "Palanpur: Population, Society, and Economy." In Lanjouw and Stern, eds.

Drèze, Jean P., and P. V. Srinivasan. 1996. "Poverty in India: Regional Estimates, 1987–88." Discussion Paper Series DEP 70. London School of Economics, Development Economics Research Programme. Processed.

Drèze, Jean P., Peter Lanjouw, and Naresh Sharma. 1998. "Economic Development 19[57]–[93]." In Lanjouw and Stern, eds.

Epstein, T. Scarlett. 1973. *South India: Yesterday, Today, and Tomorrow: Mysore Villages Re[visited].* London: Macmillan.

Frank, Robert H. 1985. *Choosing the Right Pond: Human Behavior and the Quest for Status.* [New] York: Oxford University Press.

Fuhs, F. W. 1988. *Agrarian Economy of Sunari: Stability and Change.* Saarbrucken: [...] Breitenbach.

Gadre, N. A., D. P. Wahile, and V. D. Galgalikar. 1987. "Structural Changes in Ownership [...] in a *Ryotwari* Village of Vidarabha Region." *Indian Journal of Agricultural Economics* 42[...]

Gough, Kathleen. 1981. *Rural Society in Southeast India.* Cambridge, U.K.: Cambridge Univ[ersity] Press.

———. 1987. "Socio-Economic Change in Southeast India, 1950s to 1980s." *Journal of Contemporary Asia* 17(3):276–92.

Griffin, Keith B. 1979. *The Political Economy of Agrarian Change: An Essay on the Green Revol[ution].* 2d ed. London: Macmillan.

Guhan, S., and Joan P. Mencher. 1983. "Iruvelpattu Revisited." *Economic and Political W[eekly]* (June):1063–78.

Harriss, John. 1989. "Knowing about Rural Economic Change." In Bardhan, ed.

– –. 1991. "What Happened to the Green Revolution in South India? Economic Trends, Household Mobility, and the Politics of an 'Awkward Class.'" In Hazell and Ramasamy, eds.

Harriss-White, Barbara. 1992. "Rural Poverty in India: Micro-Level Evidence." In Barbara Harriss-White, S. Guhan, and R. H. Cassen, eds., *Poverty in India, Research and Policy.* Oxford: Oxford University Press.

– – 1996. "On Long Term Village Studies." Paper presented at a dissemination workshop for the ODA funded project between Queen Elizabeth House, Oxford, and Madras Institute of Development Studies, March 27–29. Madras Institute of Development Studies. Processed.

Hazell, Peter B. R., and C. Ramasamy, eds. 1991. *The Green Revolution Reconsidered: The Impact of High Yielding Rice Varieties in South India.* Baltimore: Johns Hopkins University Press.

Heyer, Judith. 1992. "The Role of Dowries and Daughters' Marriages in the Accumulation and Distribution of Capital in a South Indian Community." *Journal of International Development* 4(4):419–36.

Iqbal, Farrukh. 1988. "The Determinants of Moneylender Interest Rates: Evidence from Rural India." *Journal of Development Studies* 24(3):364–78.

Janakarajan, V. J. 1996. "Complexities of Agrarian Markets and Agrarian Relations: A Study of Villages in Northern Tamil Nadu." Paper presented at a dissemination workshop for the ODA-funded project between Queen Elizabeth House, Oxford, and Madras Institute of Development Studies, March 27–29. Madras Institute of Development Studies. Processed.

[illegible] P. K. 1994. "Changing Conditions of Agricultural Labourers in Post-Independent India: A Case Study from Bihar." Ph.D. dissertation, Jawaharlal Nehru University, Centre for Economic Studies and Planning, School of Social Sciences, New Delhi. Processed.

[illegible]sha, N. S. 1989. "Social Science Research on Rural Change: Some Gaps." In Bardhan, ed.

Kapadia, Karin. 1993. "Mutuality and Competition: Female Landless Labour and Wage Rates in Tamil Nadu." *Journal of Peasant Studies* 20(2):296–316.

Krishnan, S. 1990. "Levels of Distortions and Inequalities in Village Systems." *The Administrator* [illegible]

Kurien, C. T. 1980. "Dynamics of Rural Transformation: A Case Study of Tamil Nadu." *Economic and Political Weekly*, Annual Number (February):365–90.

Lanjouw, Peter, and Nicholas Stern, eds. 1998. *Economic Development in Palanpur over Five Decades.* Oxford: Oxford University Press.

Leaf, Murray J. 1983. "The Green Revolution and Cultural Change in a Punjab Village, 1965–1978." *Economic Development and Cultural Change* 31(2):227–70.

Lipton, Michael, with Richard Longhurst. 1989. *New Seeds and Poor People.* Baltimore: Johns Hopkins University Press.

Mencher, Joan. 1980. "The Lessons and Non-Lessons of Kerala: Agricultural Labourers and Poverty." *Economic and Political Weekly*, Special Number (October):1781–1802.

[illegible] K. N. 1976. "Trends in Rural Unemployment in India: An Analysis with Reference to Conceptual and Measurement Problems." *Economic and Political Weekly* (August):1281–92.

[illegible]chandran, V. K. 1990. *Wage Labour and Unfreedom in Agriculture: An Indian Case Study.* Oxford: Clarendon; New York: Oxford University Press.

[illegible] Sharad. 1994. "Rural Non-Farm Employment in Uttar Pradesh, 1971–1991: A Regional Analysis." M.Phil. thesis, Centre for Development Studies, Trivandrum. Processed.

[illegible] S. M. 1972. "Land Transfers in Rural Communities: Some Findings in a Ryotwari Region." *Economic and Political Weekly* (September):A133–44.

Rodgers, G. B. 1983. "Poverty Ten Years on: Incomes and Work among the Poor of Rural Bihar. Working Paper 130. International Labour Organization, Population and Labour Policies Programme, Geneva. Processed.

Rodgers, G. B., and J. Rodgers. 1984. "Incomes and Work among the Poor of Rural Bihar [illegible] *nomic and Political Weekly.* Review of Agriculture. March.

Saith, Ashwani, and A. Tankha. 1992. "Longitudinal Analysis of Structural Change in a [illegible] Indian Village: 1970–1987." Working Paper Series 128. Institute of Social Studies, The H[illegible] Processed.

Sen, Abhijit. 1996. "Economic Reforms, Employment and Poverty: Trends and Options [illegible] *nomic and Political Weekly* (September):2459-78.

Sharma, Rita, and Thomas Poleman. 1993. *The New Economics of India's Green Revolution: In[illegible] and Employment Diffusion in Uttar Pradesh.* Ithaca: Cornell University Press.

Sreekumar, T. T. 1995. "Revisiting Village Asia: A Report on an Inter-Disciplinary Internat[illegible] Workshop." Centre for Development Studies, Thiruvanatapuram. Processed.

Srinivas, M. N. 1966. *Social Change in Modern India.* Berkeley: University of California Press.

———. 1976. *The Remembered Village.* Berkeley: University of California Press.

Srivastava, Ravi. 1995. "Beneath the Churning: Studying Development in Two Villages in [illegible] Pradesh." Paper presented at a workshop on The Village in Asia Revisited, Centre for Devel[illegible] ment Studies, January, Thiruvanatapuram. Processed.

———. 1997. "Change and Resilience in Producer Strategies in Uttar Pradesh Agriculture." In Breman, Peter Kloos, and Ashwani Saith, eds. *The Village in Asia Revisited.* Delhi: Oxford [illegible] versity Press.

Swaminathan, Madhurai. 1988. "Growth and Polarization: Changes in Wealth Inequality in [illegible] Nadu Village." *Economic and Political Weekly* (October):2229-34.

———. 1991. "Gainers and Losers: A Note on Land and Occupational Mobility in a South In[illegible] Village, 1977–85." *Development and Change* 2(2):261-78.

Unni, Jeemol. 1996. "Non Agricultural Employment and Rural Livelihoods: Macro vis-a-vis [illegible] View." *Indian Journal of Labour Economics* 39(4):795-808.

———. 1997. "Non-Agricultural Employment, Livelihoods, and Poverty in Rural India." W[illegible] ing Paper 88, Gujarat Institute of Development Research. Ahmedabad.

Visaria, Pravin, and Rakesh Basant, eds. 1994. *Non Agricultural Employment in India: [illegible] Prospects.* New Delhi: Sage.

Wadley, Susan S., and Bruce W. Derr. 1989. "Karimpur 1925-1984: Understanding Ru[illegible] through Restudies." In Bardhan, ed.

Walker, Thomas S., and James G. Ryan. 1990. *Village and Household Economies in India's [illegible] Tropics.* Baltimore: Johns Hopkins University Press.

Wiser, William, and Charlotte V. Wiser. 1971. *Behind Mud Walls.* Berkeley: University of C[illegible] nia Press.

Appraising Workfare

Martin Ravallion

rkfare programs aim to reduce poverty by providing low-wage work to those who need They are often turned to in a crisis. This article offers some simple analytical tools that be used to rapidly appraise the cost effectiveness of an existing workfare operation as a s for deciding whether the program should be expanded. For pedagogic purposes, two zed versions of a range of programs found in practice are analyzed: one for a middle-me country, the other for a low-income country. The cost of a given gain to the poor is ut the same for both programs, although the components of that cost are very different, h implications for the timing of benefits. The author points to program design changes t could enhance the impact on poverty

rkfare programs require participants to work in order to obtain benefits. Such grams have been widely used for fighting poverty in crises such as macroeco-mic or agroclimatic shocks in which large numbers of poor, able-bodied people become unemployed. Such interventions are relatively complex and difficult valuate. Other things being equal, the appraisal and design should be well in-med and rigorous. But time is short in a crisis, and data are often far from ideal. at can be done to obtain a reasonably credible and yet rapid assessment of the ly gains to the poor from a given outlay on a workfare program?

This article offers a mini-manual for the rapid appraisal of an existing workfare gram to determine if its expansion would be cost-effective. By "rapid appraisal" I n that the work can be done by two people in about two weeks with the sort of normally (though not invariably) available at short notice. Box 1 summarizes data requirements. I assume that the appraisal must address two main questions: w much impact on poverty can be expected from outlays on the existing pro-m? How might the program be modified to enhance the gains to the poor?

To illustrate, I consider two stylized versions of the programs found in practice. first is in a middle-income country in which unemployment has risen sharply in wake of a macroeconomic stabilization and reform program. The other is in a

...ld Bank Research Observer, vol. 14, no. 1 (February 1999), pp. 31–48
1999 The In...

Box 1. *Data for a Rapid Appraisal*

The rapid appraisal method proposed here requires information on

- The poverty rate in the relevant country or region
- The wage rate of unskilled (informal sector) labor
- The unemployment rate among the poor
- The labor intensity of current workfare projects, their (financial) benefit-cost ratio, their recovery rate, and the extent to which the projects are targeted to poor areas.

The likely sources of this information are household or labor force surveys, the pro[illegible] administrators (both central and local), and interviews with participants in the program.

low-income country (or region) hit by a severe drought. I label the former cou[illegible] MINC (for "middle income") and the latter LINC (for "low income").

A rapid appraisal cannot normally quantify the distribution of benefits among poor, so only the aggregate transfer from the budget is estimated here. I note, h[illegible] ever, some of the qualitative ways in which sensitivity to distribution among poor (notably by putting higher weight on gains to the poorest) can bear on appraisal and design of the program. I also note the implications of attaching [illegible] value to reducing poverty in terms of current incomes, as is often the case in a [illegible]

The program's budget is taken to be predetermined. The issues are how [illegible] effective the program is in raising the incomes of the poor, and how performa[illegible] might be improved. Such cost-effectiveness calculations can be deceptive if the b[illegible] get is not fixed. This can happen if the design of the program also affects th[illegible] sources available, by influencing how generously the nonpoor support the progr[illegible] External benefits that accrue to the nonpoor from the assets created can help m[illegible] lize broad public support. For example, insurance benefits to nonpoor particip[illegible] were a factor in public support for workfare schemes in rural areas of South [illegible] (Ravallion 1991).[2] However, it is unclear how important such considerations a[illegible] a crisis. I return to this point later.

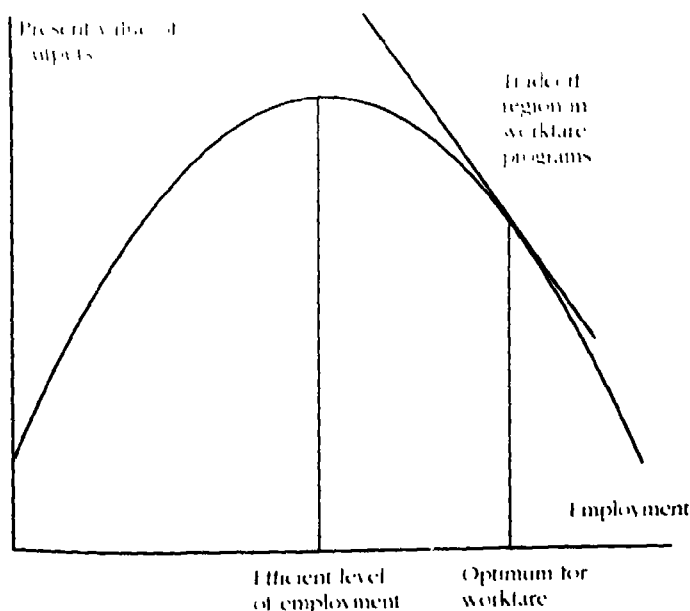
A further limitation (in common with other estimates of cost-effectiveness) [illegible] I largely ignore benefits that come in the form of better insurance. This limit[illegible] could well be serious because insurance against income losses is thought to [illegible] significant benefit from workfare programs in practice (Ravallion 1991). I do [illegible] ever, consider ways in which program design can enhance risk benefits.

Any public program must be assessed relative to the best alternative use of [illegible] same resources. The best alternative will vary with country circumstances, includ[illegible] administrative capabilities. One option that is probably feasible everywhere is a [illegible] form distribution of the program budget to every household (whether poor or [illegible] If the transfer to the poor as a percentage of total spending on the program is [illegible]

than the percentage of households that are poor, then the uniform allocation is preferable. Of course, a workfare scheme that passes this test may still be inferior to some other option; in highly industrialized countries and some transition economies, for example, a well-designed unemployment insurance scheme might be feasible and more cost-effective (Wilson and Fretwell 1996).

An Overview of the Programs

A workfare program can reduce poverty in two ways: by providing paid work for the unemployed from poor households, and by producing goods or services that poor families value. Workfare will naturally be more labor intensive than if the government simply maximized the present value of the assets created because the workfare program attaches positive value to employing poor people, independent of the gains to society as a whole from the outputs obtained. So a workfare program will tend to operate at a point where there is a tradeoff between the value of the assets created and employment (figure 1). The program will operate to the right of the point that maximizes the present value of the assets created.

Figure 1. The Tradeoff in a Workfare Program

This tradeoff poses a difficult question: How much emphasis should be given to immediate employment versus creation of durable assets? The stylized program in MINC puts relatively more emphasis on the assets created than does the LINC scheme. Municipal governments in MINC appear to use the scheme as an extra source of funds for maintaining or upgrading minor roads, sanitation facilities, and so on, using roughly the same combination of labor and nonlabor inputs as is customary. For example, a number of subprojects entailed connecting new dwellings under construction (clearly for well-to-do households) to the sewerage system. In others, pavements were repaired in well-to-do neighborhoods. In LINC the subprojects are mainly minor roads, soil conservation, reforestation, and irrigation, and the technology used tends to be somewhat more labor intensive than that used in similar projects outside the workfare program.

In neither country are the subprojects targeted to poor communities per se. In MINC the projects are just as likely to be in nonpoor neighborhoods, and in LINC the beneficiaries of rural development projects are often relatively well-off local landowners. In both cases, the work done clearly has some value to the community at large, although the projects in LINC seem unlikely to pass a conventional cost-benefit test. The municipal or provincial government usually provides cofinancing to cover the nonwage cost; local residents or nongovernmental organizations usually do not provide cofinancing. Cost recovery is rare, even from well off beneficiaries.

The Arithmetic of Cost-Effectiveness

The share of the government's outlay that benefits the poor—the cost-effectiveness ratio—can be decomposed into various components that are either estimated from the available data or calibrated from plausible assumptions. Here I suggest a decomposition that I have found useful in practice, although there are other possibilities; my aim is to provide an example that can be adapted to specific circumstances.

The cost-effectiveness ratio can be decomposed exactly into five other variables:

(i) *Budget leverage.* The government can require cofinancing from nonpoor neighborhoods for subprojects that will benefit them. Let government (central plus local) spending be G, and let this spending be leveraged up to result in a total budget $G + C$, including private cofinancing (C).

(ii) *Labor intensity.* Some of the participants may not be poor, so let the share of wages paid in total operating cost be $(W + L)/(G + C)$, where W is the wage received by the poor and L denotes leakage to the nonpoor.

(iii) *Targeted labor earnings.* This is the proportion of the wages paid out to poor workers, $W/(W + L)$.

(iv) *Net wage gain.* This is the share of the gross wage received by the poor after subtracting all costs of participation, including income forgone from other work

net wage gain is NW/W, where NW stands for wages net of forgone income or r costs of participation.

) *Indirect benefit.* Let IB denote the indirect benefits to the poor, such as when ssets created are local public goods in poor neighborhoods.

he total gain to the poor is $B = NW + IB$, which, as a proportion of public ding on the program, gives the cost-effectivness ratio:

$$\frac{B}{G} = \frac{G+C}{G} \cdot \frac{W+I}{G+C} \cdot \frac{W}{W+I} \cdot \frac{NW}{W} \left(1 + \frac{IB}{NW}\right).$$

$$\text{(i)} \quad \text{(ii)} \quad \text{(iii)} \quad \text{(iv)} \quad \text{(v)}$$

useful to further decompose the indirect benefit to net wage ratio (IB/NW) as:

$$\frac{IB}{NW} = \frac{IB}{SB} \cdot \frac{SB}{G+C} \Big/ \frac{NW}{G+C}.$$

$$\text{(vi)} \quad \text{(vii)} \quad \text{(viii)}$$

gives the net wage ratio as the product of a further three ratios:

vi) *Targeted indirect benefits.* Let the social benefits (to the whole population) the assets created be SB. The share going to the poor is then IB/SB.

ii) *Benefit-to-cost ratio for the project.* This is simply the ratio of SB to total cost, (.

iii) *The share of net wage gains in total cost.* This can also be written in terms of of the ratios in equation (1):

$$\frac{NW}{G+C} = \frac{NW}{W} \cdot \frac{W}{W+I} \cdot \frac{W+I}{G+C}$$

$$\text{(iv)} \quad \text{(iii)} \quad \text{(ii)}$$

hich the labels iv, iii, and ii correspond to the ratios in equation (1).

ome benefits, particularly the bulk of the indirect benefits, accrue in the future. can also measure the cost-effectiveness ratio in terms of current benefits (CB) by cing all values in these formulas with current values, or values within some ified period. I define the "current period" as the period during which indirect fits are negligible.

the above formulation, cost recovery from the nonpoor will increase the budget rage ratio, $(G + C)/G$, but will not change other variables. One can explicitly oduce the cost recovery rate, $k = C/(SB - IB)$, that is, the ratio of the privately nced component of the total cost to the amount of the total benefit that does not ue to the poor. One can then obtain the following formula for the budget lever- ratio:

(4) $$\frac{G+C}{G}=\left[1-k.\left(1-\frac{IB}{SB}\right)\cdot\frac{SB}{G+C}\right]^{-1}$$

(vi) (vii)

in which the labels vi and vii correspond to the ratios in equation (2).

Impact on Labor Earnings

In discussing how best to estimate plausible values for the cost-effectiveness rati[illegible] focus initially on the workfare scheme in MINC; with the basic ideas in place, [illegible] application to LINC will be straightforward. Let the data sources be the MINC Stati[illegible] tical Bureau (MSB) and the LINC Statistical Bureau (LSB).

There is a strong association between poverty and unemployment in MINC. Tab[illegible] lations from the MSB's recent national sample survey indicate an unemployment r[illegible] of 40 percent in the poorest decile of households ranked by income per pers[illegible] compared with 20 percent for all households. The unemployment rate falls stead[illegible] as income per person rises.

The Wage Rate

As is typically the case in workfare programs, no means test is applied. The sche[illegible] aims to self-select the poor and provide work for as many people as possible witho[illegible] undermining their incentive to take regular jobs when they are available. This a[illegible] proach will work if the workfare wage rate is low enough. (It should not be forgott[illegible] that the program has a budget constraint; increasing the wage rate for poor work[illegible] means that fewer poor people would benefit from the program.) A relatively [illegible] wage rate reduces the need to ration the number of workfare jobs and so enhan[illegible] the risk benefits to poor people by providing a reliable fallback in times of n[illegible] (Ravallion 1991; Ravallion, Datt, and Chaudhuri 1993).

The statutory minimum wage in MINC is $250 a month, well above the curr[illegible] monthly workfare rate of $200. (As an emergency employment program targeted [illegible] the poor, workfare is exempt from the minimum wage rate statute.) The minimu[illegible] wage, however, may also be above the market wage, given that enforcement is diffi[illegible] cult in most developing countries. So how does the workfare wage compare with th[illegible] market wages received by the poor?

In the poorest 10 percent of MINC households (ranked by household income pe[illegible] person), the average monthly earnings for the principal job (when it entailed at leas[illegible] 35 hours of work a week) were $330, well above the workfare wage. The poores[illegible]

decile received the lowest average wage among all deciles. Average wages for the second poorest and all higher deciles were more than double the workfare wage. On the basis of these data, it is reasonable to assume that the prevailing workfare wage will be unattractive to anyone who is not considered poor in MINC, and it is unlikely to attract poor workers out of their current job.

Net Wage Gain

The net wage gain (NW/W) is probably the most difficult variable to estimate in equation (1) and, possibly for this reason, it is often set to unity. This would be justified if labor supply for a workfare scheme came only from unemployment and if no other participation costs were incurred by the poor. But this assumption is difficult to accept. Poor people cannot afford to be idle. Some time will be devoted to informal, often family-run, farms or other enterprises. (This is less work than needed, but it still creates some income.) Even if all workers were unemployed at the time they joined the scheme, they would not necessarily have remained unemployed had the program not existed. Even a worker who has been unemployed for some time typically faces a positive probability of finding extra work, including self-employment or some informal sector activity. Joining the program will leave less time for search. So the net income gain will be lower than the gross wage rate paid. How much lower?

Consider a typical unemployed poor worker who is searching for a job at the time the program is opened. Without the program, the worker faces a probability P^* of finding extra work of some sort, at a wage W^*. So expected earnings without the program are P^*W^*. (One can interpret P^* as the proportion of time in which work would otherwise be found during the workfare period.)

Now introduce the workfare program. Let the probability of finding alternative work while working on the program be P (which may not be the same as P^*). The workfare wage is W. The expected income gain when the program becomes available is then $PW^* + (1 - P)W$. So the expected net wage gain (NW) to workers from introducing the scheme is:

$$NW = (1 - P)W - (P^* - P)W^*.$$

Suppose, for example, that joining the scheme means that the worker can no longer search for a regular job and hence has zero chance of getting one ($P = 0$). Then the expected gain is $W - P^*W^*$, that is, the program's wage minus expected earnings from finding a regular job. The example does not seem plausible in this setting, however. The worker can still search in nonwork hours, and participation in workfare may help in getting a regular job (by the extra experience and possibly the extra knowledge of work opportunities) sufficiently to compensate for the lost search time.

So suppose instead that joining the scheme has no effect one way or the other on the probability of finding regular work; $P^* = P$. Then the expected gain is $(1 - P)W$, that is, the proportion of the worker's time that would otherwise be unemployed times the wage rate.

In one special case, the calculation of net wage gain is greatly simplified. This occurs when no extra nonworkfare employment is available to the poor with or without the program, that is, $P^* = P = 0$. Then any income forgone by a workfare participant will be made up by an equal gain to a poor nonparticipant. Employment for the poor is then a zero-sum game. Because poor workers as a whole will forgo no income, $NW/W = 1$. It appears that this special case of zero forgone income is often (at least implicitly) assumed in discussions of workfare schemes. But zero forgone income does not seem plausible on a priori grounds, even for the poor as a whole, as discussed earlier.

What are reasonable assumptions for MINC? As already noted, the MSB survey indicated that 40 percent of those in the poorest decile were unemployed. If a worker in the poorest decile who is choosing between the program and the labor market does not accept a workfare job, he will no doubt find some work. Assuming that he has the average probability of being employed at the average wage received by workers in that decile, he will be employed 60 percent of the time at a monthly rate of $330. His expected wage if he does not accept employment through workfare is then about $200. So the current wage rate in the program of $200 turns out to be the minimum expected wage needed to attract the average worker in the poorest decile out of unemployment.

This calculation is based on averages. Actual gains to participating workers will be distributed around these averages: some workers will face relatively low chances of finding a full-time job or even casual part-time work while searching for a full-time job. Such workers will find the workfare wage rate more attractive. There are also regional differences; the same wage will be more (less) attractive in low-cost (high-cost) regions, and unemployment rates will differ from region to region. The gains from workfare will be found among those who face below-average prospects of other employment, or below-average wage rates—or both—or who live in areas where the cost of living is relatively low.

With such a high average unemployment rate in the poorest deciles, it is not unreasonable to presume that participants face unusually low prospects of finding full-time work during their spell of workfare employment. How much lower is hard to say. If program participants in MINC face a 50 percent higher unemployment rate than the poorest decile, then $P = 0.4$. Also assume that joining the program has no effect on the probability of finding a regular job. So the expected net benefit for those joining the workfare program will be 60 percent of the program wage rate. This is close to data-based estimates on the net wage ratio for an Argentinean workfare program that was implemented in 1997 (Jalan and Ravallion 1998).

he Cost-Effectiveness of the MINC Program

ne can now make a rapid appraisal of the cost-effectiveness of the workfare pro-am in MINC. Because private cofinancing is negligible, I set $C = 0$. The central ›vernment's accounts indicate that its own contributions (entirely for workfare jobs) presented one-third of total cost, so $(W + L)/(G + C) = 1/3$. Because the MINC wage te discussed earlier is very unlikely to be attractive to people who are not poor, I set = 0. From the data and assumptions discussed in the previous section, I have sumed that $NW/W = 0.6$. Then $NW/(G + C) = 0.2$.

Because MINC makes no explicit attempt to target poor areas, the poor are as likely the nonpoor to benefit indirectly from the projects; so $IB/SB = 0.2$ (the poverty te in MINC). The projects in MINC produce benefits sufficient to cover their cost; $B/(G + C) = 1$. (This reflects the fact that the labor intensity is about the average for milar public works projects.) Together, these assumptions yield $B/NW = 2$. Later I nsider alternative assumptions.

Combining these numbers, the value of B/G implied by equation (1) is 0.40. quivalently, it takes $2.50 to increase incomes of the poor by $1. Assuming that all the indirect benefit is in the future, the CB/G ratio is 0.20, so it takes $5 to transfer I to the poor today.

Because the poverty rate in MINC is 20 percent, the B/G estimate of 0.40 is double e share that poor people would obtain from a uniform, untargeted allocation of e same budget across the whole population (in which everyone gets about the same nount, whether poor or not). In terms of its impact on the current incomes of the oor, however, the workfare scheme does no better or worse than a uniform lump-im transfer to all households, whether poor or not.

he LINC Program

he LINC scheme operates primarily in rural areas. Unlike in MINC, poverty and nemployment (at least as conventionally measured) are not strongly correlated in NC. Indeed, data from the LSB indicate that unemployment rates rise as income ses, starting with the poorest, peaking at about the middle of the distribution, and alling thereafter. Substantial underemployment is thought to exist among the poor-st families, however; a worker might be classified as employed, yet work for only alf the week.

As in MINC, there is no private cost recovery ($C = 0$). The wage rate in the LINC rogram is tied to a statutory minimum wage rate for agricultural labor that is well bove the prevailing wage rate for casual unskilled agricultural labor. The high workfare age rate thus attracts participants who are not poor or unemployed. As a result, orkfare jobs are heavily rationed. Anecdotal evidence from field trips suggests that

when deciding who gets work, local program administrators do not always favor th poorest, either deliberately or because they do not know who is poor. So, unlik MINC, there is definite leakage to the nonpoor in LINC, although the forgone incom is probably lower than in MINC. I assume 0.75 for both the targeting of earning $[W/(W+L)]$ and the net wage gain (NW/W). This is consistent with an estimate o forgone income in an Indian workfare program (Datt and Ravallion 1994).

The indirect benefits to the poor are clearly smaller in LINC than in MINC; th nonpoor landowners capture the bulk of the benefits from the assets created. Th poor do receive some indirect benfits, however, notably through second-roun effects on employment from higher farm productivity. I assume that the poor ob tain one-fourth of the indirect benefits from the program. The high labor intens means, however, that the social benefits are only sufficient to cover one-half of th cost.

On plugging these numbers into equation (1), LINC's value of B/G is almost iden tical to that for MINC, 0.41, and the cost of transferring \$1 to the poor is also abo \$2.50 under LINC's program. As in MINC, it is unlikely that any of the indirect ben efits will raise current incomes (within a few months, say). The current benefit rat is 0.28 (this is CB/G, as given by the value of B/G when $IB = 0$). So it costs \$3.55 increase the current earnings of the poor by \$1 with LINC's program. Recall that th poverty rate in LINC is 50 percent. So the absolute gain to the poor from an untarget allocation of the same gross budget is *higher* than the gain from the program.

Table 1 summarizes the cost-effectiveness calculations for these two stylized pro grams under the base-case assumptions discussed above. (Costs are rounded off the nearest \$0.10.)

Table 1. *Cost-Effectiveness of the Two Workfare Programs under the Base Case Assumptions*

Variable	*Middle income country*	*Low income c*
Budget leverage: $(G+C)/G$	1 0	1 0
Labor intensity: $(W+L)/(G+C)$	0 33	0 5
Targeting: $W/(W+L)$	1 0	0 75
Net wage gain: NW/W	0.6	0 75
Poor people's share of total benefits, IB/SB	0 2	0 25
Benefit/cost ratio: $SB/(G+C)$	1 0	0 5
Current + future gains to the poor per \$1 of spending, B/G	0.40	0 41
Cost of \$1 gain to the poor (dollars)	2.50	2 50
Current earnings gain per \$1 of program spending: CB/G	0.20	0 28
Cost of \$1 extra current earnings (dollars)	5.00	3.60

Note: The poverty rate is assumed to be 20 percent in the middle-income country and 50 percent in the low income country.

Comparisons with Other Safety Net Operations

One must be cautious in comparing these estimates with cost-effectiveness ratios for other programs because the numbers are often not strictly comparable. For example, the numbers in table 1 include forgone incomes, but these are often ignored in other estimates of cost-effectiveness ratios. There are also systematic differences in the target group; for example, workfare reaches able-bodied adults, while child nutrition programs do not directly do so; rather than choosing between them, a government may need both types of program to provide a comprehensive safety net.

Assessments of safety net programs in several middle- and low-income countries show a wide range of cost-effectiveness. For cash transfer programs in Eastern Europe, Subbarao and others (1997: table 3.5) present estimates of the proportion of the public transfer going to the poor ranging from 19 to 58 percent. The same source (table 4.2) also estimates the leakage to the nonpoor from targeted food programs for several developing countries. The proportion of the total transfer to the poor ranges from 19 to 93 percent. The latter figure is an outlier; excluding it, the range is 19 to 69 percent. For food subsidy programs in India, Radhakrishna and others (1997) estimate that the share of expenditure reaching the poor is 16 to 19 percent. For housing subsidies in various countries, Subbarao and others (1997: table 4.5) estimate the share going to households below the median income at between 10 and [illegible]0 percent.

This wide range of experience makes generalizations difficult. But in terms of the impact of my stylized workfare programs on *current* incomes, one might well do worse with other instruments, particularly subsidies to goods for which demand rises with income. One could probably do better, however, or at least no worse, with an untargeted lump-sum transfer. Factoring in the estimated future income gains to the poor, the workfare programs start to look better than many other safety net operations, including untargeted lump-sum transfers, in MINC but not in HINC.

Risk Benefits

The benefits from lowering the risk of reduced income are rarely factored into calculations of cost effectiveness. How would their inclusion affect these comparisons? The risk benefits from a good workfare program can be large, as has often been demonstrated in famines (Ravallion 1997a). Even in normal times, existing (market and nonmarket) arrangements for insurance leave poor people exposed to risk (see, for example, Jalan and Ravallion, forthcoming). The risk benefits depend on the degree of risk aversion and the effect of the safety net on the riskiness of incomes, which will depend in turn on how flexibly the program responds to changing household circumstances. In this respect, some safety net programs are quite unresponsive

and therefore are as ineffective as insurance; ration cards for subsidized foods, for example, are often held for long periods and are hard to get quickly. Workfare schemes are more responsive to income risk provided the work is easily obtained when needed. That will depend on wage rates and the budget. If the wage is set so high that jobs are heavily rationed (given the available budget), the scheme will not provide reliable insurance for the poor.

This is more of a concern in LINC, yet its risk-prone rural economy is a setting in which protection is greatly needed. Insurance is difficult for the poor to obtain without incurring unacceptably high costs (including lost opportunities for escaping poverty in the longer term through potentially risky investments in human and physical capital). A full accounting of the risk benefits would probably make the MINC scheme look better but would have less effect on the calculations for LINC.

Options for Enhancing Performance

Rapid appraisals can also indicate ways in which the program's impact on poverty might be improved. Box 2 provides a checklist of recommendations for a cost-effective workfare scheme.

Box 2. *Elements of a Good Workfare Program*

- The wage rate should be no higher than the market wage for unskilled manual labor in agriculture or the informal sector during a normal year in the setting in which the program is introduced.
- Restrictions on eligibility or other forms of rationing should be avoided. Ideally, the requirement for eligibility should be the fact that work is wanted at this wage rate.
- If rationing is unavoidable (because demand for work exceeds the budget available at the wage set), then the program should be targeted to poor areas and confined to the time periods in which hardship appears to be greatest. Flexibility should be allowed in future budget allocations, however, to reflect any revealed differences in demand for the scheme.
- The labor intensity (share of wage bill in total cost) should be higher than normal for similar projects in the same setting. How much higher will depend on the relative importance attached to immediate income gains versus (income and other) gains to the poor from the assets created. The proportion will vary from setting to setting.
- The subprojects should be targeted to poor areas to ensure that the assets created are of maximum value to poor people. Any exceptions—in which the assets largely benefit the nonpoor—should require co-financing from the beneficiaries, and this money should go back into the workfare budget.
- Performance should be monitored using careful evaluation.

The Wage Rate

The wage rate for the MINC program seems about right, but the LINC wage is too high given the budget. As a result the LINC program has fewer jobs to offer than it could have, and the jobs that are available often do not go to the poor. A wage rate that is no higher than the market wage for unskilled agricultural labor in a normal agricultural year will reduce leakage, provide wider coverage of the poor with the current budget, provide better insurance, and protect incentives to take up normal work when it becomes available. It can be safely assumed that anyone who is willing to do unskilled manual labor for that wage in LINC is poor. So (assuming that the other ratios are unchanged), this alteration to the LINC program would bring the benefit ratio up to 0.50 and reduce the cost of a $1 gain to the poor to $2; LINC's workfare program would then do as well as a uniform lump-sum transfer. The current benefit ratio (CB/G) would rise to 0.375— still less than a lump-sum transfer.

Organized labor is likely to resist this reform, arguing that the government cannot undercut its own statutory minimum wage rate (although this wage is not enforceable). Labor unions in MINC initially took the same position, but the counterargument — that an exception should be made for emergency programs—won the day there.

Cost Recovery

Another way to enhance the scheme's impact on poverty is to introduce cost recovery for the benefits accruing to the nonpoor. If costs are recovered at a rate of 25 percent (but all other assumptions of the base case are retained), the value of B/G in MINC rises to 0.50, bringing the cost of transferring $1 to the poor down to $2. For LINC, $B/G = 0.45$. At a cost recovery rate of 50 percent, the value of B/G in MINC reaches 0.67, bringing the cost of an extra $1 to the poor down to $1.50. In LINC, a cost recovery rate of 50 percent is enough for B/G to reach 0.50. At a cost recovery rate of 75 percent, the cost of transferring $1 to the poor falls to $1 in MINC and to about $1.75 in LINC. Clearly, such initiatives could greatly improve program performance in reaching the poor in both countries.

Labor Intensity

Increasing the share of outlays spent on labor can greatly enhance the effectiveness of workfare programs in raising current incomes of the poor. The labor share for some projects in MINC is very low; about one-fifth are electricity and gas projects, with an average labor share of 10 percent. If these were entirely privately cofinanced, the low labor share would not be a concern. More typically, however, such projects, to be justified, would have to yield large indirect benefits to the poor. and that they should

do so seems unlikely. Data for MINC indicate that by dropping these projects overall labor share would be 0.40, implying that B/G = 0.44; this would mean th would cost $2.27 (instead of $2.50) to transfer $1 to the poor through the sche Indeed, if the MINC program had the labor intensity of the LINC program, an other characteristics were the same as in table 1, the value of B/G would rise to 0 again reducing the cost of a $1 transfer to the poor to $2.

Recall, however, that there is a tradeoff between higher labor intensity and indirect benefits from the program in both countries. As illustrated in figure workfare scheme will operate at labor intensities that entail a tradeoff betw ($W + L$) and SB. So raising the labor intensity will lower the social benefits. For sake of argument, suppose that a labor share of two-thirds in MINC was enoug drive the social benefits from the projects down to zero. Then B/NW = 1. With other ratios unchanged, the value of B/G would be 0.40, exactly what it is in current scheme. So as long as it is possible to cover at least some of the cost of scheme from the outputs generated with a labor intensity as high as that in LIN would be better to switch to a more labor-intensive scheme.

Even given the seemingly steep tradeoff with indirect benefits, a more la intensive program could be more effective in reducing poverty. The case is stronger when aiming for high current transfers to the poor.

Restrictions on Eligibility

The MINC program restricts eligibility to heads of households. By constraining family's own adjustment, however, such restrictions may actually reduce the progr effectiveness. For example, insisting that only the head of the household can join program will reduce the net gain to poor families to the extent that other househ members have less attractive labor market options and therefore lower forgone in come. The best way to raise the net transfer benefit is to let poor households arrange their own activities so as to take advantage of the workfare scheme

Designing a Program to Enhance Indirect Benefits to the Poor

Field trips to poor areas and discussions with local residents revealed plenty of su for worthwhile community infrastructure projects in both MINC and LINC. It is unreasonable to expect (pecuniary and nonpecuniary) benefits to poor people fr such projects.

To illustrate the implications for the cost-effectiveness calculations, suppose the value of the indirect benefits to poor communities from the subprojects in MI rose to half the total benefits. Keeping all other assumptions the same, B/NW wou rise to 2.25, and the overall B/G ratio would rise to 0.70. Thus the cost of transf

ring $1 to the poor would fall by more than 40 percent—from $2.50 to about $1.40. Under the same assumption, the cost of a $1 transfer to the poor in LINC would fall to $1.90. If at the same time the social benefits could be increased to cover three-fourths of the cost (still a conventional benefit-to-cost ratio under 1), a $1 gain to the poor in LINC would cost about $1.50.

Sound project selection is crucial to achieving higher indirect benefits from the assets created by workfare. Technical corroboration of the subproject's viability helps; Is it likely to work on purely technical, engineering grounds? The ex ante appraisal should, however, consider other factors likely to make the indirect benefits more pro-poor. Appraisers can identify subprojects likely to be of value to poor people and reward these using a point system, allowing subprojects to compete for funds. Census-based poverty indicators, for instance, are available at the local level in both middle- and low-income countries. Higher points should then be given to subprojects from poor areas. A municipality's past success in completing subprojects can also be rewarded with extra points. The point system could also be used to give an incentive to municipal governments in nonpoor areas that are willing to fully finance subprojects in their areas.

Direct community involvement in determining what project is to be funded and how it is to be implemented is highly desirable. This input can serve as a source of information on the likely benefits to the area and as an indicator of the subproject's longer-term sustainability. It can also help avoid "program capture" by local elites. If a bona fide local community group confirms that the subproject is valuable (even aside from the direct employment benefits), extra points should be allocated.

Other design features can enhance the value of the assets created. Relying as far as possible on workers in the same community as the project will probably improve the quality of the work because the beneficiaries have a personal longer-term interest. A bonus, contingent on successful completion to a standard that can be verified, would also help

In all cases, the central government must rely on the existing fiscal structure. Systemic factors are likely to influence the ability of some provinces to reach poor areas; for example, budget constraints can mean that poorer provinces will have a harder time targeting their own poor areas (Ravallion 1997b). Central government incentives to improve performance and provide technical assistance in proposing viable projects at the local level could help. Progress in placing subprojects in poor areas is very difficult to monitor, although this will be easier in MINC, where the statistical system is better developed.

When a points system is used and both project selection and the budget allocation are decentralized, a check should be made for horizontal inequality between areas in the minimum number of points for a subproject to be accepted. If there are large differences, then reallocations of the budget may be called for, with more

money going to areas where the minimum points needed for project acceptance are higher. A good information system for monitoring projects can help in all these respects.

Conclusion

This article shows how to conduct rapid appraisals of workfare programs. For illustrative purposes, two programs are considered, one in a middle-income country and one in a low-income country. The programs are stylized versions of those found in practice. The cost of a $1 gain to the poor using the program is about $2.50 in both cases, although the components of that cost are quite different: the poor obtain higher gains in current earnings in the low-income country, reflecting the program's higher labor intensity. The cost of a $1 gain in current earnings is $5 for the middle-income country and $3.50 for the low-income country. The amount received by the poor from a given public outlay under the program is double what they would receive from a uniform (untargeted) transfer in the middle-income country, but the poor would receive more with such a lump-sum transfer than with workfare in the low-income country. A comparison with the cost-effectiveness ratios of other types of safety net operations suggests that workfare schemes are more effective than poorly targeted food and housing subsidies but not as effective as other options. These comparisons may be deceptive, however; the same costs are not always considered, and the same options are not always feasible.

Selected reforms could enhance the benefits to the poor. It should be possible to switch to more labor-intensive production methods for subprojects in middle-income countries. There is also scope in both middle- and low-income countries for enhancing the indirect benefits to the poor from the assets created. Redistributive cofinancing—whereby cost recovery is applied only to asset creation in nonpoor areas—could also greatly improve cost-effectiveness.

Tradeoffs between some of these options are likely. In particular, too high a labor intensity will mean that the projects yield negligible indirect benefits. Circumstances will no doubt influence the choices made with respect to such a tradeoff. In a crisis situation, it is understandable that officials will opt for high labor intensity. In more normal times, where the political sustainability of the safety net is also an issue, indirect benefits will tend to get greater weight. The present calculations suggest that by any one of these routes—greater cost recovery from the nonpoor, higher labor intensity, or greater indirect benefits—design changes should make it possible to enhance appreciably the gains to the poor from a given outlay.

It cannot be denied that these calculations are rough. Naturally, the more rapid the appraisal, the more assumptions will be needed to make up for missing data. This type of appraisal is no substitute for a rigorous evaluation of a program after

has been implemented, but it can help inform public choice and program design. It can also help identify key areas where further data and analysis would have a high return.

Notes

Martin Ravallion is with the Development Research Group of the World Bank. For their helpful comments, the author thanks Harold Alderman, Polly Jones, Nora Lustig, W. R. Savedoff, K. Subbarao, Dominique van de Walle, Ann Velenchik, Michael Walton, and participants at seminars the World Bank

1. On the arguments for and against workfare programs, see Ravallion 1991; Besley and Coate 1992 Ravallion and Datt 1995; Lipton and Ravallion 1995, section 6; Mukherjee 1997; and Subbarao 1997

2. An example is the Sri Lankan food stamp scheme a better-targeted program (with the poor taking a higher share of the budget) was introduced, but this subsequently undermined political support from the middle class, and the poor ended up with less than they had before the reforms. For further discussion of these and other issues of targeting, see Besley and Kanbur (1993) and van de Walle (1998)

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems

Besley, Timothy, and Stephen Coate. 1992. "Workfare vs. Welfare: Incentive Arguments for Work Requirements in Poverty Alleviation Programs." *American Economic Review* 82(1):249–61.

Besley, Timothy, and Ravi Kanbur. 1993. "Principles of Targeting." In Michael Lipton and Jacques van der Gaag, eds. *Including the Poor.* A World Bank Regional and Sectoral Study. Washington, D.C.: World Bank

Datt, Gaurav, and Martin Ravallion. 1994. "Transfer Benefits from Public Works Employment." *Economic Journal* 104 (November) 1346–69

Jalan, Jyotsna, and Martin Ravallion. 1998. "Transfer Benefits from Workfare: A Matching Estimate for Argentina's Trabajar Program." World Bank, Development Research Group, Washington, D.C. Processed

———. Forthcoming. "Are the Poor Less Well Insured? Evidence for Rural China." *Journal of Development Economics*

Lipton, Michael, and Martin Ravallion. 1995. "Poverty and Policy." In Jere Behrman and T. N. Srinivasan, eds. *Handbook of Development Economics* Vol. 3. Amsterdam: North-Holland.

Mukherjee, Anindita. 1997. "Public Works Programmes: Some Issues." *Indian Journal of Labour Economics* 40(2):289–306.

Radhakrishna, Rokkam, Kalanidhi Subbarao, S. Indrakant, and C. Ravi. 1997. *India's Public Distribution System: A National and International Perspective.* World Bank Discussion Paper 380. Washington, D.C.

Ravallion, Martin. 1991. "Reaching the Rural Poor through Public Employment: Arguments, Evidence, and Lessons from South Asia." *The World Bank Research Observer* 6(2):153–75.

———. 1997a. "Famines and Economics." *Journal of Economic Literature* 35(September):1205–42.

———. 1997b. "Reaching Poor Areas in a Federal System." Policy Research Working Paper 1901. World Bank, Development Research Group, Washington, D.C. Processed.

Ravallion, Martin, and Gaurav Datt. 1995. "Is Targeting through a Work Requirement Efficient? Some Evidence for Rural India." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore: The Johns Hopkins University Press.

Ravallion, Martin, Gaurav Datt, and Shubham Chaudhuri. 1993. "Does Maharashtra's 'Employment Guarantee Scheme' Guarantee Employment? Effects of the 1988 Wage Increase." *Economic Development and Cultural Change* 41(January):251–75.

Subbarao, Kalanidhi. 1997. "Public Works as an Anti-Poverty Program: An Overview of Cross-Country Experience." *American Journal of Agricultural Economics* 79(May):678–83.

Subbarao, Kalanidhi, Aniruddha Bonnerjee, Jeanine Braithwaite, Soniya Carvalho, Kene Ezemenari, Carol Graham, and Alan Thompson. 1997. *Safety Net Programs and Poverty Reduction: Lessons from Cross-Country Experience.* Washington, D.C.: World Bank.

van de Walle, Dominique. 1998. "Targeting Revisited." *The World Bank Research Observer* 13(2):231–48.

Wilson, Sandra, and David Fretwell. 1996. "Public Service Employment: A Review of Programs in Selected OECD Countries and Transition Economies." Working Paper 6 on Regional Development Policies. OECD, Paris. Processed.

Public Social Spending in Africa: Do the Poor Benefit?

Florencia Castro-Leal • Julia Dayton • Lionel Demery • Kalpana Mehra

ucation and health care are basic services essential in any effort to combat poverty and e often subsidized with public funds to help achieve that purpose. This paper examines e effectiveness of public social spending on education and health care in several African ountries and finds that these programs favor not the poor, but those who are better-off. It ncludes that this targeting problem cannot be solved simply by adjusting the subsidy ogram. The constraints that prevent the poor from taking advantage of these services st also be addressed if the public subsidies are to be effective.

ublic subsidies for social services such as education and health care rest on two sic policy objectives—efficiency and equity. Efficiency gains can be achieved en the subsidies produce external benefits or correct for a market failure. Equity lso an important objective of public spending. Education and health care, in rticular, are understood to be basic services that are essential in any fight against verty. The World Bank's strategy for poverty reduction, for example, combines d based growth with human capital development (World Bank 1990). And hat public subsidies on investments that enhance human capital must benefit poor.

what extent has public social spending in Africa been effective in reaching the ? To answer this question, this article reviews the benefit incidence of govern- spending. It finds that government subsidies in education and health care are ly targeted to the poor and indeed favor those who are better-off. Improving ting to the poor involves not simply rearranging the public subsidies but also ssing the constraints that prevent the poor from accessing these services. The le examines these issues by reviewing the evidence on the benefit incidence of lth spending in seven African countries and education funding in nine countries he region.

World Bank Research Observer, vol. 14, no. 1 (February 1999), pp. 49–72

What Is Benefit Incidence?

Measuring the benefits of publicly provided goods to individuals is a matter of long-standing concern in the economics literature. For market-based goods and services the prices consumers pay can be taken as reflecting underlying values and can be used to yield measures of welfare that can be compared across individuals and over time. But when governments subsidize the provision of private goods (such as health care, education, and many infrastructure services), the supply is usually rationed and the price paid (if any) does not necessarily reflect the marginal value to individual consumers.

Two broad approaches have been pursued to measure the value to the beneficiaries of government-subsidized goods and services. The first, based on the Aaron and McGuire (1970) methodology, emphasizes the individual's own valuation of the good (that is, the demand, or virtual, price). The difficulties inherent in estimating these prices led to the development of a less demanding approach (reviewed by de Wulf 1975 and Cornes 1995) that values publicly provided goods at their marginal cost (Brennan 1976). This second approach is called *benefit incidence*; it combines the cost of providing public services with information on their use to show how the benefits of government spending are distributed across the population (Meerman 1979, Selowsky 1979, van de Walle and Nead 1995). Until recently no such studies had been undertaken in Africa. This article fills that empirical gap.

A benefit incidence analysis involves three steps:

- Estimating the unit cost per person, or unit subsidy (in current expenditures) of providing a service.[1]
- Imputing the unit subsidy to households or individuals who are identified (usually from household surveys) as users of the service. Individuals who use a subsidized public service in effect gain an *in-kind* transfer. Benefit incidence measures the distribution of this transfer across the population.
- Aggregating individuals (or households) into subgroups of the population to compare distribution of the subsidy among different groups. The most common grouping is by income or a related measure of the welfare of the individual. The studies reported here group individuals according to the total expenditure per capita of the households to which they belong.[2]

Health Spending in Africa

Recent improvements in household survey data that provide information on welfare of households and their use of public services offer an opportunity to estimate the distribution of government subsidies in the social sectors. This section

ports estimates of the benefit incidence of public health spending in Côte d'Ivoire, Ghana, Guinea, Kenya, Madagascar, South Africa, and Tanzania.

Health Care Delivery Systems

The public health systems in the seven countries are very similar (with the exception of South Africa, which has a much more developed private sector). Typically, public facilities provide more than two-thirds of the medical care in these countries. Private nonprofit (mostly charitable) organizations provide the remaining one-third. In Tanzania, for example, private nonprofit hospitals account for about half of all hospitals and about 3 percent of all health centers. Private for-profit medical care is increasing in most of the countries, but from a low base.

Table 1 shows how households respond to an injury or illness. (Country coverage varies in the tables that follow depending on the availability of data. Thus, for example, Kenya is not included in table 1, and Tanzania is not included in table 2.) These responses reflect the availability, cost, and quality of health services, as well as the circumstances of the individual households. The results are not strictly comparable across countries because the design of the survey instruments is not standardized. Moreover, the results suggest a bias: poorer households are less inclined to report illness than are their better-off counterparts. Perhaps that is because the poor accept illness as a normal feature of life and do not consider it an event. Lower reporting could also occur because poorly educated respondents are less likely to recognize untreated illnesses, a problem that is discussed by Chernikovsky and Meesook (1986) and van de Walle (1995).

Evidence from these countries shows that patterns of treatment are strikingly different across household groups:

- The poor are more inclined to self-treat than are the rich, and they are less likely to seek private modern care.
- The richest groups rely heavily on publicly provided care, particularly in Côte d'Ivoire, Guinea, and Tanzania. Only in South Africa is there evidence of the richest groups opting out of the state system in favor of private care.
- The poor rely mainly on the public system, but the private sector is important for both the poor and the nonpoor in Ghana, South Africa, and Tanzania.
- Interestingly, with the exception of Guinea, there is little reliance on traditional health providers.[3]

These countries have three-tiered public health systems, with basic clinics and dispensaries at the first level, district-level hospitals at the secondary level, and referral and specialty hospitals at the tertiary level. Resources (and hence services) are generally concentrated at the tertiary level; typically, less than 25 percent of recurrent expenditures accrue to the primary level. The public systems are traditionally

Table 1. ***Illness and Treatment Response in Selected African Countries***
(percent)

Country, year, and quintile	Ill during previous four weeks	Response of those ill			
		No care	Modern public care	Modern private care	Traditional care
Côte d'Ivoire, 1995[a]					
Poorest	30	73	26	1	—
Richest	50	35	55	10	—
Ghana, 1992[a]					
Poorest	33	59	23	14	4
Richest	58	43	28	24	5
Guinea, 1994					
Poorest	24	60	15	0	26[b]
Richest	32	31	52	6	10
Madagascar, 1993[a]					
Poorest	20	72	20	3	5
Richest	34	52	29	16	3
South Africa, 1994[a]					
Poorest	12	25	46	23	6
Richest	26	14	9	74	3
Tanzania, 1993/94					
Poorest	12	42	37	17	3
Richest	22	27	32	39	1

— Not available

a. The reference period was two weeks, so proportions were multiplied by 2 to make estimates approximately comparable.

b. Refers to private care received at home

c. Includes all other providers

Source: Côte d'Ivoire, Demery, Dayton, and Mehra (1996); Ghana, Demery and others (1995); Guinea, World Bank (1996b), Madagascar, World Bank (1996a), South Africa, Castro-Leal (1996); Tanzania, World Bank (1995).

subsidized from general revenues, although recently each country has implemented cost recovery at most public health care facilities to help finance services and improve quality. In almost all countries, health care personnel (particularly physicians) are concentrated in urban areas, where they provide tertiary-level care, and are comparatively scarce in rural areas.

Although resources and services are heavily focused on specialized health care, the main causes of illness and death in all seven countries are preventable and easily treated diseases, such as acute respiratory illness, diarrhea, and malaria. In Madagascar it is estimated that 90 percent of illnesses could be prevented or treated at the primary level, provided the services are of good quality and accessible to the majority of the population (World Bank 1996a:79). In an effort to provide better primary and preventive care, most of these countries have begun to decentralize public health

care systems. Several have recently modified the structure of their health care systems, but few have actually made major resource reallocations. In Côte d'Ivoire the share of total recurrent expenditures devoted to primary care was scheduled to increase from 35 percent in 1991 to 42 percent in 1995. Instead, the share declined in that year to 32 percent.

Unit Subsidies in Health

Estimates of the unit subsidies for public health care in six African countries are given in table 2. The unit subsidy represents the net current cost to the government of an individual visit to a health facility. It is computed as total recurrent spending on facilities, less any revenue from cost recovery (the amount that is returned to the treasury), normalized by the number of visits. Typically, this figure is obtained from government accounts. In some cases, visits are estimated from the household survey used to identify users of the facility. In others, health ministry data are used. The subsidy for a visit to a health center or primary health clinic is generally less costly to the government than a visit to a hospital, and outpatient visits are substantially less costly than inpatient visits. In Ghana an outpatient visit is one-tenth the cost of an inpatient visit, and in Guinea the ratio is 1 to 7.

Unit cost data are limited in several respects. First, only in some cases—Ghana and South Africa—do the data refer to actual recurrent spending on health facilities; in the other countries, they are based on budgeted expenditures, which may differ significantly from outcomes. Second, there is little disaggregation by type of facility, type of consultation, or region of the country, masking variations in the costs of consultation. The unit costs were generally averaged into two groups—visits to health centers and visits to hospitals. No distinction was made between different types of hospital care (such as secondary and tertiary hospitals). And making a distinction between outpatient and inpatient visits was feasible in only two countries, Ghana

Table 2. *Unit Health Subsidies by Facility in Selected African Countries*

Country	*Monetary unit*	*Health center*	*Hospital*	
			Outpatient	*Inpatient*
Côte d'Ivoire, 1995	CFAF	1,252	1,787[a]	
Ghana Accra, 1992	Cedis	6,489	4,044	49,553
Ghana other, 1992	Cedis	1,129	1,275	14,427
Guinea, 1994	GNF	902	1,321	7,926
Kenya (rural only), 1992/93	K Shs.	15	151[a]	
Madagascar, 1994	FMG	1,413	2,136[a]	
South Africa, 1992/93	DBSA	98	516[a]	

a. Average cost of all hospital visits.

Source: For Kenya, see Dayton and Demery (1994); for other countries, see table 1.

Table 3. *Benefit Incidence of Public Spending on Health for the Poorest and Richest Quintiles in Selected African Countries*
(percent)

Country/year	*Quintile shares of*								*Total subsidy as share of household expenditures*	
	Primary facilities		*Hospital outpatient*		*Hospital inpatient*		*All health*			
	Poorest	*Richest*	*Poorest*	*Richest*	*Poorest*	*Richest*	*Poorest*	*Richest*	*Poorest*	*Richest*
Côte d'Ivoire, 1995	14	22	8	39	—	—	11	32	2.0	1.3
Ghana, 1992	10	31	13	35	11	32	12	33	3.5	2.3
Guinea, 1994	10	36	1	55	—	—	4	48	—	—
Kenya (rural only), 1992	22	14	13	26	—	—	14	24	6.0	1.1
Madagascar, 1993	10	29	14	30	—	—	12	30	4.5	0.5
South Africa, 1994	18	10	15	17	—	—	16	17	28.2	1.5
Tanzania, 1992/93	18	21	11	37	20	36	17	29	—	—

— Not available

Note Hospital subsidies combine in- and outpatient spending in Côte d'Ivoire, Guinea, Kenya, Madagascar, and South Africa.

Source See tables 1 and 2

and Guinea. Further, a lack of data on regional health expenditures means that unit subsidies were generally computed at the national level. In Ghana and Madagascar, however, where regional data were available, differences among regions were significant. For example, spending per visit to a primary health care facility in Accra was almost six times that for other areas of Ghana. Such inequalities may also hold in other countries but were masked in the aggregate data to hand. It should be emphasized that the data for South Africa are for 1992–93, to correspond to the household survey year. A great deal has changed since then, with the election of the Government of National Unity. And these changes will undoubtedly influence the benefit incidence of health (and education) spending.

Who Benefits from Health Subsidies?

By combining the unit costs of health care delivery with the use of publicly funded health facilities, we can estimate the benefit incidence of government spending on health. For convenience, we report here the benefit incidence of spending to the poorest quintile (that is, the poorest 20 percent of the population, ranked by expenditure per capita) and the richest quintile (table 3).

Two clear messages emerge. First, health spending in Africa is not well targeted to the poorest. Typically the share of the subsidy to the poorest quintile was significantly less than that to the richest 20 percent. The inequality was greater in some countries (notably Côte d'Ivoire, Ghana, Guinea, and Madagascar) than in others, but overall, the poorest 20 percent of the population received less than 20 percent of the subsidy. Moreover, the share received by the richest quintile was far in excess of 20 percent (except in South Africa, where the richer households rely on private care; see table 1). The second message is that health spending is reasonably progressive; the subsidy to the poorest quintile amounts to a higher share of that group's total household expenditures than did the subsidy to the richest quintile (see table 3). This progressiveness was particularly striking in South Africa but was also true of the other countries. This finding suggests that if the government gave all households an annual income transfer rather than subsidized health care, income-expenditure distribution would improve, other things being constant.

Understanding the Benefit Incidence of Health Subsidies

To understand why health spending is not targeted to the poor in Africa, it is helpful to look at the allocation of health budgets to different levels of service—notably, hospital and nonhospital care—and the poorest quintile's share of total visits for each level of service. These two measures are obviously related: as governments change the allocations of spending across subsectors, they influence the way households choose

among treatment options, which would in all probability change the quintile shares of health visits. For convenience we examine each in turn.

The allocation of spending across services within the health sector is not favorable to the poor. Governments allocate significant shares of their health budgets to hospital based services, which the poor generally do not use. In Ghana, for example, two-thirds of the health budget was spent on hospital services; a major portion went to one large teaching hospital in Accra. In South Africa the share allocated to hospitals was [illegible] percent. And in both Madagascar and Kenya more than half of the health budget was devoted to hospitals. It is safe to say that targeting health spending to the poor in Africa would require spending less on hospitals and more on primary facilities.

Spending on hospital-based health care, however, can be justified to some extent because many large hospitals train medical personnel for lower levels of care. Moreover, one of the reasons why governments subsidize tertiary health services is that there is no insurance market. Households in developing countries cannot insure themselves against the risk of serious illness or injury and the consequent need for very expensive treatment. As the data show, this allocation of the health subsidy [illegible] be at the expense of the equity objective, because the poor tend not to use hospital services.

In Kenya, South Africa, and Tanzania, budget reallocations toward primary [illegible] would in themselves improve the targeting of spending to the poor. There, the poorest quintiles use primary facilities in good measure, gaining about one-fifth of the primary subsidy—a pattern similar to that found elsewhere in the developing world (Demery 1997). But in the other African countries, budget reallocations alone would not necessarily fix the targeting problem. In Côte d'Ivoire, Ghana, Guinea, [illegible] Madagascar the share of the subsidy received by the poorest quintile was low [illegible] levels of health care, including primary facilities. Given the costs and benefits [illegible] volved, household decisions about using publicly subsidized health care services [illegible] sult in far fewer visits to primary facilities from poor households than from better [illegible] ones. The point is that budget reallocations must be accompanied by increased [illegible] of primary facilities by poor households. To identify the interventions that [illegible] have this effect, it is necessary to understand why the poor limit their use of pub[illegible] funded health facilities.

We consider here five principal factors that affect the use of health services [illegible] poor: income, service quality, access, direct user charges, and gender.

INCOME. Health care is a normal good, which means that household spending [illegible] health—and the use of health facilities—increases with income (table 4). [illegible] table 1 shows, the richest groups use mainly publicly subsidized health care [illegible] in South Africa). In Ghana the richest quintile directed almost 60 percent of [illegible] health spending to the public sector, much of it on hospital consultations (Demery and others 1995). This means that health spending is very unlikely to be targeted [illegible]

poor. Given the fundamental influence of income on the demand for health care, only way in which public subsidies can be well targeted to the poorest is by rting the demand for health care by those who are better-off to the private sector. t is no easy task in countries where private health care is generally poorly devel-d, largely because of the dominance of the public sector. This change has to onsidered a long-term objective. In the meantime, are there other factors ame-le to shorter-term policy interventions that might mitigate this powerful income ct?

LITY DIFFERENCES. Alderman and Lavy (1996) report that the demand for health is sensitive to the quality of the service provided. Even the poor limit their and for health care when services are of poor quality. But the poor are less sensi-to *changes* in quality of service (Lavy and Germain 1994). Thus uniformly poor-lity service would discourage demand more among the rich than the poor, which ld be inconsistent with the observed share of each quintile's participation in lth care services. The observed pattern can therefore be explained only by signifi-t *differences* in the quality of service offered to the rich and the poor. So, for nce, drug availability, staff skills, and the quality of health facilities may vary siderably and to the disadvantage of the poor. Is there any evidence of such varia-s in quality? A special survey of health facilities designed to accompany the Ghana ing Standards Survey of 1989 suggests that there is (Lavy and Germain, 1994:13). d the earlier discussion of unit subsidies in Ghana implied the presence of large iations in quality (table 2). Similarly, in Antananarivo, the region in Madagascar re most (23 percent) of the poor live, the government unit subsidy for basic

ble 4 *Per Capita Household Spending on Health in Selected African Countries*

[illegible]	Health spending	Share of health spending (percent) Nonfood expenditure	Total expenditure
Côte d'Ivoire, 1988			
Poorest	3,347	13.4	5.4
Richest	14,107	6.3	3.7
Ghana, 1992			
Poorest	2,964	12.7	4.6
Richest	12,452	7.5	3.4
Madagascar, 1993/94			
Poorest	1,133	6.9	1.8
Richest	4,581	1.5	0.7

Spending figures for Côte d'Ivoire are in CFAF, for Ghana in cedis; and for Madagascar in FMG.
Source: See table 1.

health care was just 41 percent of the subsidy going to the richest region, Antsiranana, with a total poverty headcount of only 7 percent (World Bank 1996a). These comparisons suggest that there may well be differences in the care provided at different health facilities, to the disadvantage of poorer households.

ACCESS AND OPPORTUNITY COSTS. Poor households, which are often some distance from government health facilities, typically face long journeys and high opportunity costs to obtain health care. In South Africa, for example, those in the poorest quintile must travel almost two hours on average to obtain medical attention, compared with an average of 34 minutes for the richest quintile (Castro-Leal 1996a). The Ghana Living Standards Survey of 1992 also recorded longer travel and treatment times for poorer households. Time spent away from economic activity represents much greater private opportunity costs for the poor, who, unlike their salaried counterparts, have to forgo income in order to obtain medical care. These costs can dominate the decision to seek care.

Lavy and Germain (1994) found that halving the distance to public health facilities in Ghana would increase their use among the population at large by an estimated 96 percent. In Kenya distance was also a significant factor in the demand for health care, although not as dramatic as in Ghana (Mwabu, Ainsworth, and Nyamete 1993). Lavy and Germain (1994) found that the poor were willing to pay less than the nonpoor in absolute terms, but more relative to their income, to reduce the distance traveled. Gertler and van der Gaag (1990) found that individuals at the lower end of the income distribution in Côte d'Ivoire were far more sensitive to changes in the time required to obtain care than were those at the upper end. Time, in effect, rations the market. These studies are based on cross-sectional evidence, however; the direction of causation is uncertain; the relationship between use and distance may be capturing the effect of geographic variations in health care utilization on government decisions about placement of health facilities rather than the other way around.

PRICE. The cost of a medical consultation is far more of a burden for the poor, and ample evidence suggests that when prices are raised through cost recovery schemes, the poor are more likely than the nonpoor to cut back on their use of health services (Gertler and van der Gaag 1990; Lavy and Germain 1994). Longitudinal studies based on controlled experiments such as those by Litvack and Bodart (1993) in Cameroon and by Gertler and Molyneaux (1997) in Indonesia confirm that price increases without compensating improvements in quality discourage utilization by the poor. Increasing user charges, other things being equal, lowers the share of the poor in total visits to health facilities. Charges must therefore be introduced carefully; they must be targeted to services used mainly by the nonpoor; and if applied to services used by the poor, they should be accompanied by improvements in both access and quality.

ble 5. *Benefit Incidence of Health Spending by Gender for Selected African Countries*

Country/year	Quintile	Percentage share	
		Female	Male
Côte d'Ivoire, 1995	Poorest	52.9	47.1
	Richest	55.0	45.0
Ghana, 1992	Poorest	44.3	55.7
	Richest	65.0	35.0

Source: See table 1.

ENDER. Income, quality, access, and price interact with social relationships to pro-
uce sharp inequalities in the distribution of health benefits by gender. Females in
e top quintiles in Côte d'Ivoire and Ghana, for example, typically use publicly
inded health facilities more than do their male counterparts (table 5). But this
dvantage changes markedly for the poorest quintiles. The gender advantage is largely
oded in Côte d'Ivoire, although poor females still use facilities more than do males,
nd in Ghana poor females use health facilities less than do males in the same quintile.
he reasons for this are unclear. Supply-side effects may account for the difference—
e facilities available to the poor may not provide the perinatal care used by their
cher counterparts. Or the difference may arise from demand-side household pref-
ences. Poor households may decide that females should not use health facilities,
ither because of underlying social values favoring males over females or because of
igher opportunity costs of female time. Either way, poor households behave differ-
ntly from rich households, and this difference explains to some extent the weak
argeting of the health subsidy to the poor in Africa.

ducation Spending in Africa

ducation has long attracted government subsidies in Africa, in part because of the
xpected high social externalities involved, but also because of equity considerations.
he case for subsidizing primary education is particularly strong, given the wide
nefits it brings. Literacy and numeracy are critical to sustaining modern democra-
es. And a growing weight of evidence from the endogenous growth literature high-
ghts the favorable growth effects of education (Bruno, Ravallion, and Squire 1996;
emery, Sen, and Vishwanath 1995).

haracteristics of the Education Systems

ormal education in all nine countries for which we have data (Côte d'Ivoire, Ghana,
uinea, Kenya, Madagascar, Malawi, South Africa, Tanzania, and Uganda) includes

six years of primary school (seven in Tanzania), three years of lower secondary school, three years of senior secondary school, and four years of university education. Most public systems have vocational, technical, and teacher-training programs parallel to the university system. Movement through the educational system is generally determined by student performance in national examinations. The government is the main provider of education in all nine countries, although the size of the private sector varies substantially. At one extreme, for-profit provision of primary education is prohibited in Tanzania (although the number of private secondary schools is increasing dramatically). And at the other, almost 30 percent of primary and secondary students in Accra attend private schools.

Public education is financed by both governments and households. Of total current spending on education in Ghana, the government contributes about 65 percent and households about 35 percent (Demery and others 1995). Household out-of-pocket contributions include school fees, uniforms, books, supplies, and the like. Households also incur opportunity costs (of the time forgone while attending school) as well as transaction costs (mainly transportation to and from school). Attendance fees vary: in Côte d'Ivoire and Tanzania primary schooling is free, but in Ghana and Guinea nominal fees are charged at all levels.

Although all governments consider primary education to be the highest priority, the degree to which the budget reflects this priority varies. The share of the education budget allocated to primary schooling ranged from just 40 percent in Guinea in 1994 to more than 70 percent in Malawi in 1994–95. Wages and salaries dominate the functional categories in the budget. In Ghana and Malawi wages accounted for 94 and 9[illegible] percent, respectively, of total costs, and in Tanzania the share was 81 percent. Elsewhere the distribution between salaries and supplies was not as skewed, with salary expenditures accounting for between one-half and two-thirds of recurrent expenditures.

Enrollment rates vary by education level and household income in the following ways (table 6):

- Enrollment rates in primary schools are generally lower than the average for low-income countries, although variation among them is substantial.
- Enrollment rates are extremely low at the secondary level, at around 1[illegible] percent, substantially lower than the average for low-income countries worldwide (42 percent for girls and 55 percent for boys in 1993, according to World Bank 1996c). The exception is South Africa, where secondary education is almost universal.
- Enrollment rates are significantly lower for the poor at all levels, and particularly at the secondary level. Again, South Africa is an exception, with both primary and secondary rates among the poorest quintile being close to the national average.
- The *overall* enrollment rates for boys and girls at the primary level is about the same in many of these countries, but a gap emerges among poorer quintiles.

ble 6. *Gross Enrollment Rates in Primary and Secondary Education for the Poorest d Richest Quintiles in Selected African Countries*

(cent)

ntry/year	Primary			Secondary		
	Poorest	Richest	All	Poorest	Richest	All
te d'Ivoire, 1995	51	99	75	12	65	31
ana, 1992	75	101	88	27	45	39
inea, 1994	16	84	44	2	40	19
nya, 1992[a]	100	108	105	9	55	31
dagascar, 1993	48	113	83	2	53	18
lawi, 1994, 1990[b]	74	133	108	4	30	10
uth Africa, 1994	112	97	106	81	101	97
nzania, 1993/94[c]	[illegible]	87	81	3	20	10
anda, 1992	72	116	93	4	43	19

Decile averages.

Primary enrollment rates are for 1994 and secondary enrollment rates for 1990.

Unweighted average of male and female enrollment rates.

rces: Côte d'Ivoire, Demery, Dayton, and Mehra (1996); Ghana, Demery and others (1995); Guinea, ld Bank (1996b); Kenya, Demery and Verghis (1994); Madagascar, World Bank (1996a); Malawi, Castro- (1996); South Africa, Castro-Leal (1996); Tanzania, World Bank (1995), and Uganda, Ablo and Reinikka 8).

large gender gap in enrollments is evident at the secondary level for most income groups.

Low enrollment rates are not the only indicator of poor performance; repetition s are uniformly high — more than 30 percent — and completion rates are corre- ondingly very low. In addition, most of the nine countries have a problem with starting age. In Tanzania more than 80 percent of all primary school students late in enrolling: the average starting age was 9 for girls and 10 for boys (Mason Khandker 1997: 5).

nit Subsidies in Education

nit subsidies for education are computed as net recurrent spending (total govern- ent recurrent spending less cost recovery to the treasury) per student. In most of studies reported here, unit subsidies are based on government expenditure data nd enrollment estimates from household surveys; in some cases, tertiary enroll- ents are based on government statistics. For most countries the unit subsidies ap- y to the country as a whole, taking into account only differences between the levels education. But for Madagascar and South Africa, it is possible to disaggregate rther (box 1). Unit subsidies increase with the level of education, markedly so in uinea and Malawi (table 7). Typically the outlays for secondary schools are about

Box 1. *Disaggregating Unit Subsidies and Education Spending*

Where spending is unevenly distributed geographically (or in other ways), the use of aggregate ***unit subsidies can mask inequality in public spending.*** In South Africa Castro-Leal (1996b) obtained five levels of unit subsidy based on the budgets of the different "houses" of government, which were divided along racial lines. Unit subsidies varied enormously, but enrollment rates were high, even among the poorest groups receiving the lowest subsidy. In Madagascar unit subsidies were obtained for the six main regions of the country (World Bank 1996b). The subsidies did not vary as much as in South Africa, but enrollment rates declined sharply at low income levels.

Two estimates of the benefit incidence of education spending are reported in the table below. One is based on the disaggregated unit subsidies, while the other is computed using an average unit subsidy at each of the three education levels. In South Africa the disaggregation of unit subsidies makes a significant difference to benefit incidence. For education spending as a whole, the use of average subsidies makes it appear as though each quintile received roughly its proportionate share of the education budget. But in fact, the poorest quintile gained only 14 percent and the richest [illegible] percent of total education spending because of unit cost variations between the races. The Madagascar estimates tell a quite different story. Here, the use of national average unit subsidies (at each level of schooling) changes the benefit incidence estimates only marginally compared with the use of [illegible]-specific unit subsidies. The differences are literally a matter of tenths of a percentage point.

Why the difference between South Africa and Madagascar? Three factors explain this outcome. First, the unit subsidies were far more variable in South Africa than they were in Madagascar, [illegible] as they did the years of the apartheid regime. Second, the population within the quintiles [illegible] distributed across regions in Madagascar, so that there was some variability in the unit subsidies within quintiles. In South Africa the population in the poorest quintile was almost entirely [illegible], and it was the black population that received the lowest unit subsidy. Third, enrollment rates [illegible] uniformly high in South Africa, whereas enrollment rates in Madagascar varied significantly across income groups. The lower enrollment rates among the poorer groups in Madagascar were [illegible] caused in part by the lower unit subsidies allocated to them. When national average unit subsidies are used, variations in the unit subsidy are missed, but their effects on the enrollment patterns [illegible] income are captured and are reflected to some extent in the benefit incidence estimates.

Two Measures of Benefit Incidence of Education Spending in South Africa and Madagascar
(percent)

Population quintile	Primary spending: Disaggregated	Primary spending: Mean	Secondary spending: Disaggregated	Secondary spending: Mean	Tertiary spending: Disaggregated	Tertiary spending: Mean	Education [illegible]: Disaggregated [illegible]
	Share of quintile in total subsidy						
South Africa, 1994							
Poorest	18.9	25.8	11.5	18.8	6.1	6.1	[illegible]
Richest	27.8	13.5	38.6	16.6	47.2	[illegible]	[illegible]
Madagascar, 1993							
Poorest	16.8	17.2	1.9	2.0	0.0	0.0	8.2
Richest	14.4	14.0	41.8	41.5	88.6	88.6	41.2

Source: Castro-Leal (1996b), World Bank (1996a)

Table 7. *Unit Education Subsidies by Level in Selected African Countries*

Country/year	Monetary unit	Primary	Secondary	Secondary as ratio of primary	Tertiary	Tertiary as ratio of primary
Côte d'Ivoire, 1995	CFAF	64,840	117,462	1.8	348,453	5.4
Ghana, 1992	Cedis	24,824	65,275	2.6	392,707	15.8
Guinea, 1994	GNF	47,625	116,812	2.5	2,595,705	54.5
Kenya, 1992/93	K Shs.	1,368	3,868	2.8	42,050	30.7
Madagascar, 1994	FMG	50,504	192,491	3.8	1,140,000	22.6
Malawi, 1994/95	Kwachas	220	909	4.1	15,523	70.6
South Africa, 1994	DBSA	1,124	2,055	1.8	5,657	5.0
Tanzania, 1993/94	T Shs.	6,600	7,500	1.1	—	—
Uganda, 1992/93	U Shs	11,66[illegible]	3[illegible]352	3.2	373,525	32.0

— Not available.

Source: See table 6.

[...]ice the amount spent on primary schools. Tertiary unit subsidies were significantly greater than other levels.

Who Benefits from Education Subsidies?

Combining the unit cost data with information on the use of publicly subsidized education from household surveys yields estimates of the benefit incidence of government education spending. The subsidy for education, like that for health, is generally progressive but poorly targeted (table 8). In absolute terms, the poorest quintile gets less than 20 percent of the subsidy — significantly less in most cases (Côte d'Ivoire, Guinea, Madagascar, South Africa, Tanzania, and Uganda). The richest quintile gets far more, especially in those same five countries. Yet the subsidy for public education is more equally distributed than household income or expenditure. The monetary benefit to the poor, as a share of total household expenditure, is more than the benefit to the rich, particularly in Kenya and South Africa. Generally education subsidies represent a greater gain to poor households in these countries than do health subsidies.

The high share of the primary school subsidy imputed to the poorest quintile, shown in table 8, is misleading because the education *needs* of this group are so much greater than those of other groups. In both Côte d'Ivoire and Ghana, the share of [...] school age children in the poorest quintile is much greater than the share of [...] that quintile receives (table 9). The contrast between needs and benefits is [...] more striking in the case of secondary school subsidies. The poorest quintile in Côte d'Ivoire accounts for 21 percent of secondary-school-age children but receives [...] percent of the subsidy.

Table 8. *Benefit Incidence of Public Spending on Education in Selected African Countries*
(percent)

Country/year	Quintile shares of total spending								Total subsidy as share of household expenditures	
	Primary subsidy		Secondary subsidy		Tertiary subsidy		Total subsidy			
	Poorest	Richest	Poorest	Richest	Poorest	Richest	Poorest	Richest	Poorest	Richest
Côte d'Ivoire, 1995	19	14	7	37	12	71	13	35	12.5	4.6
Ghana, 1992	22	14	15	19	6	45	16	21	13.4	3.1
Guinea, 1994	11	21	4	39	1	65	5	44	—	—
Kenya, 1992	22	15	7	30	2	44	17	21	27.8	1.9
Malawi, 1994	20	16	9	40	1	59	16	25	2.3	1.4
Madagascar, 1993	17	14	2	41	0	89	8	41	7.2	3.4
South Africa, 1994	29	28	11	39	6	47	14	35	42.1	5.1
Tanzania, 1993/94	20	29	8	34	0	100	14	37	—	—
Uganda, 1992	29	18	4	49	6	47	13	32	4.3	1.5

— Not available.
Source: See table 6.

Table 9. *Benefit Incidence and Education Needs in Côte d'Ivoire and Ghana*
(percent)

Country/quintile	Primary: Share of subsidy	Primary: Share of school-age population	Secondary: Share of subsidy	Secondary: Share of school-age population
Côte d'Ivoire				
Poorest	19.1	23.8	6.8	20.9
Richest	13.9	13.5	37.2	20.9
Ghana				
Poorest	21.8	24.3	14.9	20.4
Richest	14.1	13.7	18.6	16.8

Source: See table 6.

These demographic differences across the quintiles arise in part because of the use of per capita household expenditures as the welfare measure; as a result of that measure, poor households are both larger and have more children than better-off households. Because of this, they gain a significant proportion of the primary education subsidy. If, instead, *per adult equivalent* expenditures were used, these demographic differences might disappear (Lanjouw and Ravallion 1994). To see whether their outcomes were sensitive to the measures of welfare used, Demery and others (1995) normalized household expenditures on both household size and adult equivalence (using a scale proposed in Deaton and Muellbauer 1986) and found that spending was significantly less targeted to the poorest under the revised welfare measure and that a larger share went to the richest quintile (table 10). The exercise confirmed the sensitivity of the benefit incidence results to the welfare measure (see also van de Walle, Ravallion, and Gautam 1994, Jarvis and Micklewright 1995).

Table 10. *Benefit Incidence of Education Subsidies under Alternative Welfare Measures, [illegible] 1992*
([illegible])

[illegible]	[illegible]	Per capita expenditures
Primary		
Poorest	1[illegible]	21.8
Richest	19	14.0
Secondary		
Poorest	18.6	14.9
Richest	16.6	18.6
Tertiary		
Poorest	9.5	6.0
Richest	29.8	45.2

Source: Demery and others (1995).

Unlike health, the share of the education subsidy accruing to the poorest varie ticeably by level of service. The poorest quintile is seen to gain far more from sp ing at the primary level—typically about one-fifth of the subsidy, compared about one-tenth of the subsidy at the secondary level and almost nothing fror subsidy at the tertiary level. Thus the more governments spend on primary ec tion, the more the poor will benefit.

Yet, understanding the differential enrollment rates remains important if go ments are to improve the targeting of education subsidies to the poor. Althou rich and growing literature exists on the constraints facing the poor in acce health services, far less evidence is available on the demand for education in A Again, the major determinants of demand are income, quality, and costs (oppo nity and direct costs).

Table 11 shows how much the poorest and richest income groups in three c tries spend on education. Typically, rich households spend more than the p Unless better-off groups can be encouraged to use private service providers, e cially at the secondary and tertiary levels, it is difficult to envisage how governm education subsidies can be better targeted to the poor.

Education systems in most of the countries reviewed here need to be impro Less well documented is variation of service quality within a country and the ex to which the poor are disadvantaged. Unit cost variations suggest that the ser provided in poorer rural communities are inferior to those extended to urban h

Table 11. *Household Spending on Education by Poorest and Richest Quintiles in Selected African Countries*

Country/quintile	Year	Education spending per capita	Education spending as percent of: Nonfood expenditures	[illegible] expenditures
Côte d'Ivoire	1995			
Poorest		2,083	5.1	[illegible]
Richest		23,964	6.9	[illegible]
Ghana	1992			
Poorest		1,924	8.3	[illegible]
Richest		6,872	4.2	[illegible]
Madagascar	1993/94			
Poorest		338	2.0	[illegible]
Richest		3,000	1.0	[illegible]

a. Spending figures for Côte d'Ivoire are in CFAF, for Ghana in cedis, and for Madagascar in [illegible]

Source: See table 6.

hools. Case and Deaton (1998) reported that pupil-teacher ratios in black schools ere more than twice those in white schools under the former regime in South Af- ca, and their econometric results suggest that this policy discouraged school enroll- ent and educational attainment among black households. Evidence from Uganda ows that the amount that actually reaches rural schools is much less than aggregate ta would suggest. Ablo and Reinikka (1998) found that for each dollar the govern- ent spent on primary education, only 36 cents actually reached the rural schools. his means lower-quality education in such areas.

In Ghana Lavy (1992) found that leaking, unusable classrooms and lack of elec- icity had significantly negative effects on decisions to enroll in primary schools, id Glewwe and Jacoby (1992) reported that other variables (for example, no desks) so influenced primary enrollment. The government has increasingly required local mmunities to meet capital and nonwage recurrent costs, which means that educa- nal services in poor communities are likely to be inferior because their resources e so constrained.

Households that enroll children in school encounter costs, such as the costs of pplies and transportation and the loss of the children's work. What effect do these sts have on education? Mason and Khandker (1997) could find little evidence in nzania that out-of-pocket costs had a negative effect on enrollments. But work on hana by Lavy (1992), by Norton and others (1995), and more recently by Chao d Alper (1998) suggests that these costs do reduce enrollments. Tan, Lee, and lingat (1984) estimated the elasticity of school enrollment with respect to changes i direct costs at about -0.5 in Malawi. But all these studies failed to establish whether e poor are more sensitive to these costs than the population at large, thus explain- ig the observed pattern of school enrollment across income groups. Intuition would ggest that they are, but evidence from Africa is not available. (For evidence from donesia, see King 1995, on Peru, see Gertler and Glewwe 1989.)

The distance to the nearest school might also explain why enrollments are so low nong poorer households. For example, Lavy (1992) found distance an important rminant in Ghana. But interestingly, it is not the distance to the primary school at influences primary school enrollments, but the distance to postprimary school- . Parents are clearly making decisions based on the whole education investment ycle. The decision to enroll girls was more sensitive to the access costs of postprimary cation than was the decision to enroll boys. Because distances to secondary schools longer than those to primary schools for poorer households, enrollments are y to be low at all levels. Although primary schools are more widespread and ssible, access is still a problem for many poor rural communities in Ghana: Chao d Alper (1998) found that enrollment was almost 70 percent in communities with ast one primary school in 1992, compared with only 43 percent in those with no mary school. They estimated that reducing the distance to a primary school by e mile would raise enrollment by 3 percentage points. In contrast, Mason and

Khandker (1997) report that distance is important only in secondary school enro-
ment in Tanzania, largely because most communities in that country are served
primary schools.

In most of the countries covered here, fewer girls than boys from low-inco
groups attend primary school, and fewer girls than boys from all income grou
attend secondary school. This means that gender bias is an important explanati
for the poor targeting of education spending in Africa (see Demery 1997).

Concluding Observations

This review does not present a particularly encouraging picture. Although spendi
on social services is usually justified on equity and efficiency grounds, most heal
and education subsidies in the region are not particularly well targeted to the poo
est. Subsidies to primary education are an exception, but even here, they appe
inequitable when judged against the numbers of school-age children in the poo
groups and when alternative measures of economic welfare are used. There are groun
for considering that the inequality results shown here underestimate the true i
equality. Regional variations in unit subsidies and in the quality of services pro
vided—largely ignored in the results reported here—are likely to further disadva
tage the poor. Moreover, the poor are less able than the better-off to augme
government subsidies by contributing to the services obtained (table 12).

One of the most fundamental factors responsible for weak targeting is the pos[illegible]
income elasticity of demand for these services. In the long run, the strategy mu[illegible]
to encourage private providers so that the public subsidy can be directed more e[illegible]
tively at services used mostly by the poor. But there are instruments that could [illegible]
prove targeting in the short and medium term. The first involves reallocating pu[illegible]

Table 12. *Social Sector Spending on Poorest and Richest Quintiles in Ghana, 1992*

	Government subsidy		Household spending[a]		[illegible]
Sector/quintile	*Mean*	*Share*	*Mean*	*Share*	*Mean*
Health					
Poorest	2,296	12	1,998	10	[illegible],294
Richest	6,515	33	[illegible],099	37	13,614
Education					
Poorest	8,731	16	1,761	10	10,492
Richest	11,067	21	5,072	30	16,139

a. Household spending on publicly subsidized services.

Note: Mean is in cedis per capita; share is percentage of spending for all quintiles.

Source: Demery and others (1995).

subsidies toward services used primarily by the poor. On this, the evidence is mixed: in health, many of the poor do not use any services very much, even primary services. Expenditure reallocations would improve targeting only if they led to a significant increase in the use of such services by the poor. In education, there is somewhat more scope to channel resources to the poor through primary education, but even here, enrollment rates are low, especially among girls.

Changes in household behavior, therefore, are critical. Two factors appear to be important: quality of service and access to facilities. The poor are not well served by the public provision of health services. Such services that are available are costly to access. The evidence suggests that the poor would be willing to increase their use of health services if both quality and access could be improved. It also points to the need for increased attention to the infrequent use of health services by poor women. Improving quality and reducing cost would also seem to be critical for raising the demand for education among poor communities in Africa, although here the evidence is somewhat patchy.

A well-designed user-fee policy could potentially improve the benefit incidence of health and education spending, but the decision to impose such fees should be undertaken with care. Fees should be applied to services where *total* demand (for private and public services) is price inelastic and where good-quality private services are available. They should not be applied where good-quality private services are unavailable or where the demand is very price elastic (those services used mainly by the poor, for example). If user fees are combined with significant improvements in both access and quality, there is growing evidence that the poor will increase their use of the service.

Benefit incidence has provided important insights into the problems facing governments in Africa that are struggling to deliver essential social services to poor communities. But although it highlights the problems, it is short on answers. For Africa, at least, the message is that reallocations of public expenditures are not sufficient; policies must be based on a sound understanding of the factors that govern household decisions about health care and schooling and of the means by which subsidized services can lead to better outcomes for the poor.

Notes

1. [illegible] Castro-Leal is an economist and Kalpana Mehra is a research analyst in the Poverty Reduction and Economic Management Network at the World Bank. Julia Dayton is a Ph.D. candidate in health economics at Yale University, and Lionel Demery is lead poverty specialist in the Poverty Reduction and Social Development Group in the Africa Regional Office of the World Bank.

2. Current expenditures are used because they benefit *current* beneficiaries. Capital spending may have a quite different incidence, but it benefits *future* beneficiaries.

2. This welfare measure, now an established one for poverty analysis, is described in Ravall (1993). Usually, household expenditures include imputed values for own-produced consumpti and take into account regional variations in prices. In most cases, the welfare measure normali household expenditure on household size (the exception here being the study of South Africa, wh uses total household expenditure per adult equivalent as its welfare measure). As shown here, results are sensitive to the welfare measure used. Our reliance on per capita expenditure comes fr the use of this measure in the studies that we draw on. But for future work, benefit incidence sho explore ranking households by other measures. The effects of different assumptions about eco mies of scale in household consumption, for example, should be investigated. Given the effect random variations in observed expenditures, an alternative approach would be to use instrumen or predicted values of the welfare measure (see Behrman and Knowles 1997).

3. These data are not comparable across countries. The Guinea estimate, for example, assu that all private treatment provided at the home of the respondent is traditional (which is not tru the other countries). There is some suggestion in these data that respondents report visits to pri pharmacists and traditional caregivers as "self-treatment," which would explain the apparently use of traditional care. Such underreporting of traditional care would leave reported use of mod health providers (and the analysis that follows) unaffected. There may well be *indirect* effe government health spending that affect such services, but without strong empirical evidence ab these effects in the studies reviewed here, we assume that such care is unaffected by public subsi

References

The word "processed" describes informally reproduced works that may not be commonly av through library systems

Ablo, Emmanuel, and Ritva Reinikka. 1998. "Do Budgets Really Matter? Evidence from P Spending on Education and Health in Uganda." Policy Research Working Paper 1926. W Bank, Africa Region, Washington, D.C. Processed.

Aaron, Henry, and Martin C. McGuire. 1970. "Public Goods and Income Distribution." *Econom* 38(6):907–20.

Alderman, Harold, and Victor Lavy. 1996. "Household Responses to Public Health Service and Quality Tradeoffs." *The World Bank Research Observer* 11(1):3–22.

Behrman, Jere R., and James C. Knowles. 1997. "How Strongly Is Child Schooling Associated Household Income?" University of Pennsylvania, Philadelphia, Pa. Processed.

Brennan, Geoffrey. 1976. "The Distributional Implications of Public Goods." *Econ* 44:391–99.

Bruno, Michael, Martin Ravallion, and Lyn Squire. 1996. "Equity and Growth in Developme tries: Old and New Perspectives on the Policy Issues." Policy Research Working Paper World Bank, Office of the Vice President, Development Economics and Policy Resea partment, Washington, D.C. Processed.

Case, Anne, and Angus Deaton. 1998. "Large Cash Transfers to the Elderly in South Afr *nomic Journal* 108(September):1330–61.

Castro-Leal, Florencia. 1996a. "The Impact of Public Health Spending on Poverty and Ine in South Africa." PSP Discussion Paper Series 101. World Bank, Poverty and Social P partment, Washington, D.C. Processed.

———. 1996b. "Poverty and Inequality in the Distribution of Public Education Spending Africa." PSP Discussion Paper Series 102. World Bank, Poverty and Social Policy Depar Washington, D.C. Processed.

———. 1996c. ***Who Benefits from Public Education Spending in Malawi: Results from the Recent Education Reform.*** World Bank Discussion Paper 350. Washington, D.C.: World Bank.

ao, Shiyan, and Omer Alper. 1998. "Accessing Basic Education in Ghana." Studies in Human Development, Number 1. World Bank Human Development III, Africa Region, Washington, D.C. Processed.

ernikovsky, Dov, and Oey Astra Meesook. 1986. "Utilization of Health Services in Indonesia." *Social Science and Medicine* 23(6):611–20.

rnes, Richard. 1995. "Measuring the Distributional Impact of Public Goods." In van de Walle and Nead, eds.

yton, Julia, and Lionel Demery. 1994. "Public Health Expenditure and the Rural Poor in Kenya." World Bank, Education and Social Policy Department, Washington, D.C. Processed.

aton, Angus, and John Muellbauer. 1986. "On Measuring Child Costs: With Applications to Poor Countries." *Journal of Political Economy* 94(4):720–44.

Wulf, Luc. 1975. "Fiscal Incidence Studies in Developing Countries." *IMF Staff Papers* 22:61–131.

mery, Lionel. 1997. "Benefit Incidence Analysis." World Bank, Poverty Reduction and Economic Management Network, Washington, D.C. Processed.

mery, Lionel, and Mathew Verghis. 1994. "The Incidence of Education Expenditure in Kenya." World Bank, Education and Social Policy Department, Washington, D.C. Processed.

mery, Lionel, Julia Dayton, and Kalpana Mehra. 1996. "The Incidence of Social Spending in Côte d'Ivoire, 1986–95." World Bank, Poverty and Social Policy Department, Washington, D.C. Processed.

mery, Lionel, Binayak Sen, and Tara Vishwanath. 1995. "Poverty, Inequality, and Growth." ESP Discussion Paper Series 70. World Bank, Education and Social Policy Department, Washington, D.C. Processed.

emery, Lionel, Shiyan Chao, Rene Bernier, and Kalpana Mehra. 1995. "The Incidence of Social Spending in Ghana." ESP Discussion Paper Series 82. World Bank, Poverty and Social Policy Department, Washington, D.C. Processed.

ger, Paul, and Paul Glewwe. 1989. *The Willingness to Pay for Education in Developing Countries: Evidence from Rural Peru.* LSMS Working Paper 54. Washington, D.C.: World Bank.

ger, Paul, and John Molyneaux. 199_. "Experimental Evidence on the Effect of Raising User Fees for Publicly Delivered Health Care Services: Utilization, Health Outcomes, and Private Provider Response." Rand Corp., Santa Monica, Calif. Processed.

ger, Paul, and Jacques van der Gaag. 1990. *The Willingness to Pay for Medical Care: Evidence from Two Developing Countries.* Baltimore: Johns Hopkins University Press.

we, Paul, and Hanan Jacoby. 1992. *Estimating the Determinants of Cognitive Achievement in* [illegible] *Income Countries: The Case of Ghana.* LSMS Working Paper 91. Washington, D.C.: World Bank.

Sarah J., and John Micklewright. 1995. "The Targeting of [illegible] Allowances in Hungary." In van de Walle and Nead, eds.

Elizabeth. 1995. "Does the Price of Schooling Matter? Fees, Opportunity Costs, and Enrollment in Indonesia." World Bank, Development Research Group, Washington, D.C. Processed.

ow, Peter, and Martin Ravallion. 1994. "Poverty and Household Size." Policy Research Working Paper 1332. World Bank, Policy Research Department, Washington, D.C. Processed.

Victor. 1992. *Investment in Human Capital: Schooling Supply Constraints in Rural Ghana.* LSMS Working Paper 93. Washington, D.C.: World Bank.

Lavy, Victor, and Jean-Marc Germain. 1994. *Quality and Cost in Health Care Choice in Developing Countries.* LSMS Working Paper 105. Washington D.C.: World Bank.

Litvack, Jennie, and Claude Bodart. 1993. "User Fees Plus Quality Equals Improved Access to Health Care: Results of a Field Experiment in Cameroon." *Social Science and Medicine* 37(3):369–83.

Mason, Andrew, and Shahidur Khandker. 1997. "Household Schooling Decisions in Tanzania." World Bank, Gender Anchor Group, Poverty Reduction and Economic Management Network, Washington, D.C. Processed.

Meerman, Jacob. 1979. *Public Expenditures in Malaysia: Who Benefits and Why?* New York: Oxford University Press.

Mwabu, Germano, Martha Ainsworth, and Andrew Nyamete. 1993. "Quality of Medical Care and Choice of Medical Treatment in Kenya: An Empirical Analysis." *Journal of Human Resources* 28(4):838–62.

Norton, Andy, Ellen Bortei-Doku Aryeetey, David Korboe, and D. K. Tony Dogbe. 1995. "Poverty Assessment in Ghana Using Qualitative and Participatory Research Methods." PSP Discussion Paper Series 83. World Bank, Poverty and Social Policy Department, Washington D.C. Processed.

Ravallion, Martin. 1994. *Poverty Comparisons.* Vol. 56, *Fundamentals of Pure and Applied Economics.* Chur, Switzerland: Harwood Academic Publishers.

Selowsky, Marcelo. 1979. *Who Benefits from Government Expenditure?* New York: Oxford University Press.

Tan, Jee-Peng, Kiong Hock Lee, and Alain Mingat. 1984. *User Charges for Education: The Ability and Willingness to Pay in Malawi.* World Bank Staff Working Paper 661. Washington D.C.

van de Walle, Dominique. 1995. "The Distribution of Subsidies through Public Health Services in Indonesia." In van de Walle and Nead, eds.

van de Walle, Dominique, and Kimberly Nead, eds. 1995. *Public Spending and the Poor: Theory and Evidence.* Baltimore: Johns Hopkins University Press.

van de Walle, Dominique, Martin Ravallion, and Madhur Gautam. 1995. "Testing a Social Safety Net." *Journal of Public Economics* 57(2):175–99.

World Bank. 1990. *World Development Report 1990: Poverty.* New York: Oxford University Press.

———. 1995. "Tanzania Social Sector Review." Report 14039-TA. Washington D.C. Processed.

———. 1996a. "Madagascar Poverty Assessment. Vol. II: Main Report Annexes." Report [illegible]-MAG. South-Central and Indian Ocean Department, Africa Regional Office, Washington D.C. Processed.

———. 1996b. "Republic of Guinea Public Expenditure Review." Report [illegible] West Africa Department, Washington D.C. Processed.

———. 1996c. *World Development Report 1996: From Plan to Market.* New York: Oxford University Press.

Rethinking the Causes of Deforestation: Lessons from Economic Models

Arild Angelsen • David Kaimowitz

This article, which synthesizes the results of more than 140 economic models analyzing the causes of tropical deforestation, raises significant doubts about many conventional hypotheses in the debate about deforestation. More roads, higher agricultural prices, lower wages, and a shortage of off-farm employment generally lead to more deforestation. How technical change, agricultural input prices, household income levels, and tenure security affect deforestation—if at all—is unknown. The role of macroeconomic factors such as population growth, poverty reduction, national income, economic growth, and foreign debt is also ambiguous. This review, however, finds that policy reforms included in current economic liberalization and adjustment efforts may increase the pressure on forests. Although the boom in deforestation modeling has yielded new insights, weak methodology and poor-quality data make the results of many models questionable.

Concern is rising about the adverse consequences of tropical deforestation. The loss of forest cover influences the climate and contributes to a loss of biodiversity. Reduced timber supplies, siltation, flooding, and soil degradation affect economic activity and threaten the livelihoods and cultural integrity of forest dependent people. Tropical rain forests, which constitute about 41 percent of the total tropical forest area, are considered the richest and most valuable ecosystem on the earth's land surface. During the 1980s about 15.4 million hectares of tropical forests were lost each year, according to estimates by the United Nations Food and Agriculture Organization (FAO 1992). From 1990 to 1995 the annual loss was estimated at 12.7 million hectares (FAO 1997), but it is unclear whether this reduction represents a slowdown in actual forest clearance or new definitions and better data.

This concern has led economists to expand their efforts to model why, where, and to what extent forests are being converted to other land uses. Kaimowitz and Angelsen

The World Bank Research Observer, vol. 14, no. 1 (February 1999), pp. 73–98

(1998), in a comprehensive review of more than 140 models, describe why la
holders behave the way they do and examine the links between the larger econo
and decisions to clear—or to protect—the forest. The models vary with regard to
precise definition of forest, if indeed they provide any definition at all. In most
stances in this paper, the term *deforestation* describes the complete long-term
moval of tree cover. Like all social science models, those discussed here simp
complex multidimensional processes and highlight only a few of the many varial
and causal relations involved in changing patterns of land use. These models, ho
ever, do allow one to think about deforestation more systematically and to expl
the possible effects of policy or other exogenous changes on land use.

A Framework for Analyzing Deforestation

The conceptual framework used here is helpful both in understanding deforesta
processes and in classifying modeling approaches. Five types of variables are used
models of deforestation:

- *The magnitude and location of deforestation* the main dependent variable
- *The agents of deforestation*—those individuals, households, or companie
  volved in land use change and their characteristics
- *The choice variables* those decisions about land allocation that determine
  overall level of deforestation for the particular agent or group of agents
- *Agents' decision parameters* those variables that directly influence agents
  cisions but are external to them
- *The macroeconomic variables and policy instruments* —those variables that a
  forest clearing indirectly through their influence on the decision parameters

Figure 1 illustrates the relations among the main types of variables and provide
simple, logical approach to analyzing deforestation at three different levels sour
immediate causes, and underlying causes. This schematic varies somewhat from
existing literature, which is rather inconsistent in its use of these terms. The star
point is to identify the agents of deforestation (small farmers, ranchers, logge p
tation companies) and their relative importance. These agents' actions are the
of deforestation. Theoretically at least, the magnitude of various sources can be
rectly measured—although it may be difficult to do so and no economic analy
required.

Next one might focus on agents' decisions, which are based on their own char
teristics (background, preferences, and resources) and on decision parameters s
as prices, technology, institutions, new information, and access to services and in
structure. Together, these factors determine the set of available choices and the

Figure 1. *Variables Affecting Deforestation*

Underlying causes of deforestation

Macroeconomic-level variables and policy instruments

↓

Immediate causes of deforestation

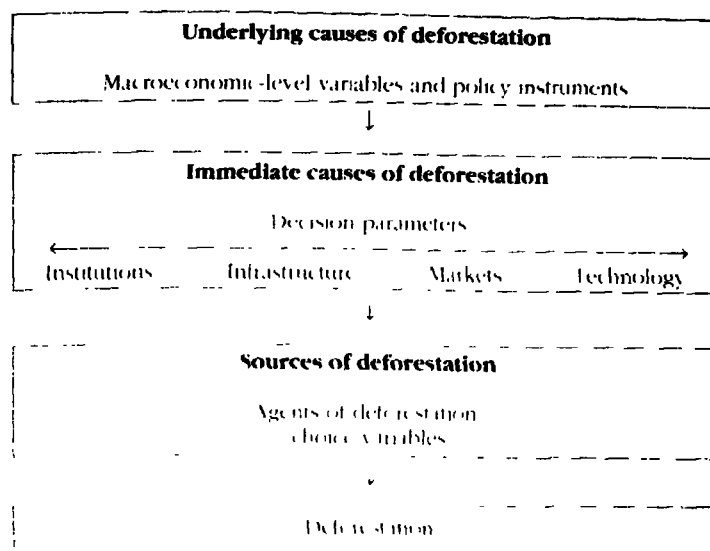
Decision parameters

← Institutions — Infrastructure — Markets — Technology →

↓

Sources of deforestation

Agents of deforestation
choice variables

↓

Deforestation

Source: Authors' construction.

entives for different choices. The decision parameters may be seen as the ***immediate*** *uses* of deforestation.

Finally, the agents' characteristics and decision parameters are themselves deter-mined by broader forces. These *underlying causes* of deforestation influence agents' cisions through several channels—the market; the dissemination of new technolo-es and information; the development of infrastructure; and institutions, particu-arly the property regime.

For the sake of simplicity, figure 1 and the discussion so far imply that causal lations go in only one direction. But important effects also go in the opposite rection, for example, the decisions agents make will have important feedback ef-ects on market prices (general equilibrium effects). Agents' collective actions, politi-al pressures, and demographic behavior also affect underlying causes.

A clear distinction among the three levels is necessary for several reasons. First, it s useful to single out the parameters that directly affect decisionmakers. Second, the evels of variables are related to the type of model used: microeconomic models focus n immediate causes, whereas macroeconomic models tend to deal with underlying auses. Third, because the underlying causes determine the immediate causes, which n turn influence the agents who are the sources of deforestation, mixing these levels

confuses the causal relations involved and leads to serious misspecifications in regression models. And fourth, the results regarding the sources and immediate causes are more conclusive than those for the underlying causes.

Although the article focuses on the immediate and underlying causes of deforestation, some comments on the sources are in order. There is a broad consensus that the expansion of cropped areas and pastures is a major source of deforestation and that the expansion of pastures is especially important in Latin America (Kant and Redantz 1997). No similar consensus has formed about logging, although it seems to be a direct source of deforestation in some contexts and an indirect source in others. Logging roads, for example, facilitate access to the forest by farmers (Burgess [illegible]). Southeast Asia is one region in which logging has contributed significantly to deforestation. Evidence regarding fuelwood collection and open-pit mining is weak, though it implicates them as the sources of some significant deforestation, particularly for fuelwood in Africa.

Surprisingly little is known about how the characteristics of agents affect behavior. Researchers know that subsistence-type households are less responsive to market signals than families who are more market oriented, but existing models [illegible] little about the prevalence of such behavior. Nothing significant can be generalized from available information about the role of farm size, farmer background, or [illegible] company characteristics. The conventional poverty environment argument is that poorer families are more likely to clear the forest, either to grow crops or [illegible] wood, because they have shorter time horizons (higher discount rates). [illegible] counterargument says such families are less likely to do so because they lack [illegible] necessary capital to put additional land into production (see, for example, [illegible] 1993). Existing models provide little evidence on this issue.

Analytical and empirical models suggest that time preferences and risk aversion are important to farmers and loggers. But their practical effect depends on what relevant investment decisions are assumed to be. High discount rates and risk aversion are both likely to reduce investment, but that investment could be either to [illegible] the forest or to conserve it (Southgate 1990; Mendelsohn 1994; von Amsberg [illegible]). These two types of investment are not symmetrical, however: whereas forest clearing requires action, forest conservation requires little more than leaving the forest [illegible]. This suggests that forest clearing is more likely to be considered the relevant investment decision.

The Models Reviewed

We have reviewed more than 140 papers containing economic models that represent key processes associated with deforestation. The exclusive focus on formal models does not imply that these models are necessarily more useful or more accurate [illegible]

informal studies based solely on descriptive statistics. Such "soft" analyses and studies complement formal models and offer important insights that are difficult to capture in formal models.

Quantitative models have many clear limitations. They focus on variables for which quantified data are available. They do not, typically, explicitly address issues related to market failure (users do not capture the full value of preserving tropical forests), which one could argue are the "real" underlying causes of deforestation (Pearce 1996). Further, institutional factors are rarely included. One could argue, however, that the variables related to market failure or institutional arrangements are fairly stable over time compared with prices, for example, and are therefore less relevant to changes in rates of deforestation.

Global reviews such as this one inevitably emphasize the similarities between countries and regions, rather than their differences. The factors affecting deforestation, the interactions between them, and the magnitude of their effects all vary significantly from one location to another. Models based on data from distinct locations can reach conflicting conclusions not only because they use distinct definitions, variables, or methodologies, but also because the processes themselves differ.

In this context, it should be noted that most of the empirical and simulation models that analyze a single country or region focus on a handful of countries: Brazil, Cameroon, Costa Rica, Indonesia, Mexico, Thailand, and, to a lesser extent, Ecuador, the Philippines, and Tanzania. Most of these countries are medium or large in size and population, are relatively politically stable, and have large areas of tropical rain forest, so the results presented here may be more applicable to countries with those characteristics.

Categories of Models Reviewed

Table 1 shows the distribution of the models analyzed here. We have classified the models based on two criteria: scale (household and firm, or microeconomic, level;

Table 1 *Models of Deforestation by Category*

	Analytical	*Simulation including programming*	*Regression*	*Total*
Household and firm	15	9	9	33
Regional	0	3	30	33
National	19[a]	23	38	80
Total	34	35	77	146

a. This figure includes an impressive 14 papers by Jones and O'Neill. Most have a similar methodological framework, but each has distinct features and addresses different issues.

Source: Authors' calculations.

regional level; and national, or macroeconomic, level) and methodology—analytical, simulation, and empirical.[1]

Analytical models are abstract, theoretical constructs. They include no empirical data but rather clarify the implications of different assumptions about how agents behave and how the economy operates, which may not be obvious. Simulation models use parameters based on stylized facts drawn from various sources to assess scenarios. Most simulation models at the microeconomic level are whole-farm analyses using (linear) programming techniques, whereas the most common macroeconomic simulation models are computable general equilibrium (CGE) models. Empirical models quantify the relations between variables based on empirical data. Almost all empirical models use regression analysis, usually the standard ordinary least squares (OLS) method. Below we describe the three types of model.

Microeconomic Models

As the name suggests, these models seek to explain how individuals allocate their resources, using standard economic variables such as background and preferences, prices, institutions, access to infrastructure and services, and technological alternatives. A major distinction is between models that assume all prices are market determined and farmers are fully integrated into perfect markets (Southgate 1990; Mendelsohn 1994; Bluffstone 1995; Angelsen 1999) and those that do not (Dvorak 1992; Holden 1993; Angelsen 1999). Within the former category, production decisions are guided by market prices (including off-farm wages) and can be studied as a profit-maximizing problem. When farmers are not fully market integrated, decisions are based, in part, on farmers' subjective (and endogenously determined) shadow prices. Factors such as resource endowments (poverty) and household composition are important, and the consumption side must be included when making the production decision. This distinction turns out to be critical for how model makers predict land use will change in response to changes in population, agricultural prices, and income. Analytical models have been very useful in highlighting the role played by the underlying market and behavioral assumptions (Angelsen 1999).

The main strength of farm-level simulation (programming) and regression models reviewed here lies in their use of generally good-quality survey data regarding the magnitude of deforestation and description of farmers' behavior. Strictly speaking, however, these conclusions apply only to the area studied. Some of the conclusions of the simulation models depend heavily on the market assumptions discussed above, which the models normally do not test. Farm-level regression models often say little about farmers' response to price changes because the price variation within the area is normally too small to allow such analysis.

Regional Models

The coverage in such models is limited to a region or area with a distinct and characteristic ecology, agrarian structure, institutional and political history, set of trade networks, and pattern of settlement and land use (Lambin 1994:16). Analytical or simulation models rarely focus on a region, although there are a few exceptions.[2] Although deforestation is inherently a spatial phenomenon, most models lack an explicit spatial dimension; thus they cannot answer the *where* question.

Most regional models are thus regression models, which may be spatial or nonspatial. Spatial models measure the impact on land use of variables such as how far the forest is from markets and roads, topography, soil quality, precipitation, population density, and zoning categories. This type of analysis has become more popular since the advent of digitized land use data and geographic information systems that have made it easier to manipulate the data. Nonspatial models, however, are more common. These models use data obtained at a provincial or regional level in a manner similar to multicountry regression models, but the regional models generally have better data on forest cover: about half the models reviewed here used satellite data, either alone or in combination with land surveys.

Decisions affecting the rate of deforestation are taken at the household level, but the most interesting consequences affecting biodiversity and watersheds often occur at the district or regional level. Accounting for behavioral changes of farmers and other agents is difficult in spatial models. It should soon be possible in some cases to use panel data for spatial regression models, which will facilitate the inclusion of price variables (Foster, Rosenzweig, and Behrman 1997). It should also be possible to incorporate agricultural census and survey data into a geographic information systems framework, which would allow modelers to take into account many additional variables.

Macroeconomic Models

National and multicountry models emphasize the relations among underlying variables, decision parameters, and deforestation. Analytical, simulation, and regression models are all well represented at this level.

To model complex macroeconomic processes in a strictly analytical framework and still reach interesting conclusions, model makers have generally had to place strict limits on the number of variables and make some strong assumptions. Both analytical and computable general equilibrium (simulation) models at the national level add two important dimensions to the analysis that are absent in household- and firm-level models. First, they make some prices endogenous. Thus they move beyond simply asking how decision parameters influence agents and look at how the underlying variables determine one particular set of decision parameters (prices).

This provides an important link to macroeconomic variables and policy instruments. Second, most models include the interactions among different sectors, for example (subsectors of) agriculture, forestry, and manufacturing, which makes them useful in analyzing the underlying causes of deforestation.

Some computable general equilibrium (CGE) models take a conventional approach and assume that land is a factor of production and that forest is cleared up to the point where the current land rent is zero (Coxhead and Jayasuriya 1994; Aune and others 1997). Others pay particular attention to the property regime (Persson and Munasinghe 1995; Unemo 1995). A third group applies a forest rotation (Faustmann) approach (Thiele and Wiebelt 1994; Thiele 1995). CGE models can be criticized for the poor quality of their data and the parameters commonly used, their questionable assumptions about perfect markets, and (particularly in the case of the forest rotation approach) their descriptions of farmers' or loggers' behavior. In such models [illegible] conclusions depend heavily on the responsiveness of the variables to changes in prices and income, and these elasticities are often chosen rather arbitrarily.

Multicountry (global) regression models *comprise the single largest category* [illegible] deforestation models. They rely on national data to make global generalizations [illegible] the major processes affecting tropical deforestation. But problems with the methods and the data make their usefulness and validity questionable. First, most researchers use deforestation data from the Food and Agriculture Organization Forest Resources Assessments (1981, 1992) or from the FAO production yearbooks. We agree [illegible] Rudel and Roper (1997: 54) that neither is "acceptable for empirical analysis of causes of deforestation" because they are based largely on dubious data sources or mere extrapolations based on forest cover data from a single point in time. For example, in the 1990 assessment (FAO 1992), only 21 of the estimates for the 90 countries were based on two or more national forestry inventories. For the remaining countries, deforestation rates were extrapolated from a single data point using a model with population density and ecological classes as its only explanatory variables. [illegible] countries had no forest inventory at all: of the 66 countries with one inventory [illegible] inventories were taken before 1981. The data for African countries are particularly poor.

Because of the difficulty of obtaining reliable data, many multicountry regression models use the percentage of forest land as a proxy for deforestation. Kummer [illegible] Sham (1994) argue persuasively, however, that forest cover depends on the percentage of land originally in forests and the total amount of forest cleared through human history and is not related in any simple way to recent deforestation. Moreover, many models mix sources, immediate causes, and underlying causes in [illegible] independent variables. (The work of Kant and Redantz 1997 is an exception to [illegible] general picture.) Besides potential statistical problems of multicollinearity and [illegible] ased estimates, this mixing will also distort the interpretation of cause and effect.

And finally, to produce meaningful cross-country results, it is important that the ariables included affect deforestation in roughly the same manner across countries. This is obviously a strong assumption because studies indicate that the effect of and nteraction among economic growth, foreign debt, population, and other variables may differ greatly from one country to the next. In principle, this problem could be vercome by adding interaction terms among the independent variables, but in practice, the degrees of freedom are too small to do that.[3]

In sum, most of the existing multicountry regression models do not accurately stimate the direction and size of the effects that different variables have on deforestation. Because of these weaknesses, we have given less weight to these models in the discussion.

The Immediate Causes of Deforestation

The main source of deforestation is clearing by households or companies for agriculture or timber. The question is, what factors make farmers and loggers decide to clear more forests? Table 2 gives an overview of the main results of the models.

Agricultural Prices

Substantial evidence supports the assertion that higher prices for agricultural products stimulate forest clearing. As frontier agriculture becomes more profitable, both the existing population and migrants from other areas begin to shift resources into forest clearing. Higher prices also provide capital to put additional land into agricultural production.

At the theoretical level, there is only one reason why higher agricultural prices might not increase deforestation: when farmers exhibit a preference for subsistence-farming, they will opt for leisure once they have reached some minimal consumption level. In this case they will produce less when prices are higher because they can meet their basic consumption needs without clearing more land. Microeconomic simulation models that assume subsistence behavior, such as Ruben, Kruseman, and Hengsdijk (1994) and Angelsen (1999) find less deforestation when agricultural prices are higher, while models that assume farmers are profit maximizers show the opposite (Monela 1995).

Although it is possible that some households might respond to higher agricultural prices by reducing the amount of land farmed, there is no evidence for this at more aggregated levels. Regional regression models on Mexico by Barbier and Burgess and Deininger and Minten (forthcoming), on Sudan by Elnagheeb and Bromley (1994), on Tanzania by Angelsen, Shitindi, and Aarrestad (1998), and on

Variable	Effect of increase in variable, by model type: Analytical	Simulation and empirical	Comments
Agricultural output prices	Increase	Increase	Farm-level analytical models predict increase, unless there are strong income effects (subsistence models)
Agricultural input prices	Indeterminate	Mixed	Fertilizer price increases may induce shift to more land-extensive systems
Off-farm wages and employment	Reduce	Reduce	Among the most significant findings
Credit availability	Indeterminate	Increase[a]	Depends on whether the relevant investment is forest clearing or forest management and agricultural intensification; most studies find that credit finances deforestation
Technological progress on frontier farms (direct effects)	Indeterminate	Little evidence	Similar to price increase; new labor-intensive technologies may reduce deforestation if labor supply is inelastic.
Accessibility (roads)	Increase	Increase	Among the most significant findings, although roads are partly endogenous
Homesteading property regime	Increase	Little evidence	Claims to future land rents give farmers an additional incentive to clear land
Land tenure security	Indeterminate	Increase[a]	Empirical evidence is relatively weak
Timber prices	Indeterminate	Increase[a]	Empirical findings are weak but tend to find a positive link

a. Data may not be reliable.
Source: Authors' analysis.

Thailand by Panayotou and Sungsuwan (1994) all find a positive correlation between higher agricultural prices and deforestation. Binswanger and others (19 found a positive correlation between total cropped area and agricultural prices in cross-country analysis of 58 countries. All the analytical macroeconomic and computable general equilibrium models also show that increased agricultural prices boost deforestation, although this result is as much a product of their initial assumption that farmers are profit maximizers as it is of the empirical evidence.

It should be emphasized that this discussion refers only to changes in the aggregate terms of trade for agriculture with respect to other sectors. Changes that affect the relative prices of different crops and livestock products may have quite different effects. Thus it is impossible to predict how specific policies will affect forest clearing

ithout looking at their impact on prices for specific products and the pressure each roduct puts on forests. For example, Gockowski (1997) shows that deforestation creased in Cameroon after relative prices shifted in favor of plantains, the production of which requires substantial forest clearing, from cocoa, which requires less and.

Although the prices of agricultural products and other decision parameters can be ken as given by the individual farmer, they are not truly exogenous in the models as is also the case for many of the other variables in table 2). Output prices are a unction of total supply. The regression models reviewed do not attempt to separate out predetermined (exogenous) changes (taxes, exchange rates, and so on) from the esponse to these changes. The response to an exogenous price increase will dampen he initial increase, but this effect is likely to be small because output from recently cared land often has a small market share.

rices of Agricultural Inputs and Credit

e theory of how changes in agricultural input prices affect forest clearing leads to determinate conclusions, and the empirical evidence is mixed, particularly for ferzers. Analytical models point to two conflicting effects. On the one hand, higher ilizer prices lead farmers to adopt more extensive production systems that use re land and less fertilizer. On the other hand, the higher costs associated with reased fertilizers make agriculture in general less profitable and can lead to a reation in the amount of land devoted to crops.

Attempts to resolve the issue empirically have been only partially successful. Linprogramming and regression models suggest that fertilizer price increases in south- Africa provoke greater deforestation or have little impact (Monela 1995; Aune others 1997; Holden 1997; Mwanawina and Sankhayan 1996), whereas in some in American contexts such price increases may reduce deforestation (Barbier and gess 1996). Higher fertilizer prices seem most likely to induce greater forest clearing en farmers are wavering between intensive sedentary agriculture and more extenshifting cultivation systems. This finding adds a cautionary note about the posle negative impact of current policies aimed at reducing fertilizer subsidies in Subaran Africa (Holden 1997).

The evidence regarding the prices of other agricultural inputs, such as seeds, pesides, and hand tools, suggests that higher prices reduce forest clearing (Ruben, seman, and Hengsdijk 1994; Ozorio de Almeida and Campari 1995; Monela 95). In these cases the reduced profitability of agriculture appears to outweigh any ift toward more extensive production.

In theory, credit expansion could reduce the pressure on forests if it were used more intensive agriculture or for forest management investments. It will, howr, increase the pressure if used to finance activities associated with forest clear-

ing, such as extensive cattle ranching. Most empirical evidence on credit comes from farm- and regional-level regression analysis in tropical Latin America and concludes that credit availability is positively correlated with deforestation (Ozorio de Almeida and Campari 1995; Barbier and Burgess 1996; Andersen 1997; Pfaff 1997). The only significant exceptions are two studies of indigenous farmers in Bolivia and Honduras, which found that farmers who used credit deforested less (Godoy and others 1996, 1997). In these cases, families with credit may be less dependent on forest-based activities or may choose to engage in off-farm work to repay their loans. Modeling work in Africa and Asia has largely ignored the issue of credit availability (with the exception of Monela 1995, who finds a positive relationship between credit availability and forest clearing in Tanzania), perhaps because it is less important there.

Wages and Off-Farm Employment

All types of microeconomic models strongly suggest that higher rural wages reduce deforestation by making agricultural and forestry activities more costly. They also suggest that, at the individual household level, greater off-farm employment opportunities produce a similar effect by competing with such activities for labor (Holden 1993; Ruben, Kruseman, and Hengsdijk 1994; Bluffstone 1995; Godoy and others 1996, 1997; Pichón 1997).

Regional and national analytical and simulation models also support these conclusions, although the hypotheses have yet to be successfully validated in macroeconomic empirical models because of limited data on wages and off-farm labor. One has, therefore, strong reasons to believe that policies that favor rural wage increases and generate off-farm employment opportunities for rural people should reduce deforestation. Such policies should simultaneously conserve forests and diminish poverty.

Technological Progress in Agriculture

Technology has both a direct effect on farmers' behavior and an indirect effect resulting from its impact on product and factor prices (including wages). We focus here on the first set of effects, leaving the second for a later section.

Technological changes that increase yields without significantly altering labor or capital requirements can be expected to increase deforestation. The extent of forest clearance is likely to be even greater if technological changes are labor- or capital-saving, or both, since this will free up resources for farming additional land (Southgate 1990). Conversely, if the new technology is more labor- or capital-intensive and farmers find it difficult, expensive, or inconvenient to hire wage labor or obtain credit, then such changes can lead farmers to devote more labor and capital to their

isting farms, leaving them with fewer resources for expansion. Under these cir-
umstances the net effect is indeterminate (Larson 1991). More generally, technolo-
ies that make more intensive production systems more profitable reduce the need
or clearing additional forest land for agriculture, according to linear programming
odels by Nghiep (1986) and Holden (1993).

These findings imply that agricultural research and extension policies designed to
mit deforestation should focus on promoting profitable technologies that are labor-
nd capital-intensive and more easily applicable to land already under cultivation.
he empirical evidence, however, is still limited, and this is clearly an important area
or future research.

ccessibility and Roads

nalytical and empirical models and studies find that greater access to forests and
arkets accelerates deforestation. Roads, rivers, and railroads all facilitate access.
orest fragments are more accessible than large compact forests, and forests in coastal
ountries and islands are more accessible than those in continental countries (Krutilla,
yde, and Barnes 1995; Rudel and Roper 1996).

Spatial regression models are well suited for studying the effects of access. Models
of this type for Belize (Chomitz and Gray 1996), Cameroon (Mertens and Lambin
997), Costa Rica (Sader and Joyce 1988; Rosero Bixby and Palloni 1996), Hondu-
as (Ludeke, Maggio, and Reid 1990), Mexico (Nelson and Hellerstein 1997), and
he Philippines (Liu, Iverson, and Brown 1993) all show a strong relation between
oads and deforestation. Several find a similar result between proximity to markets
nd forest edges. Most studies show that forest clearing declines rapidly beyond dis-
ances of 2 or 3 kilometers from a road, although Liu, Iverson, and Brown (1993)
eport significant forest clearing up to around 15 kilometers from the nearest road.
These results are also supported by nonspatial regression models from Brazil (Andersen
1997; Pfaff 1997), Ecuador (Southgate, Sierra, and Brown 1991), the Philippines
Kummer and Sham 1994), and Thailand (Panayotou and Sungsuwan 1994; Crop-
per, Griffiths, and Mani 1997).

The simple correlation between distance to roads and deforestation found in re-
gression models tends to overstate the causality, since some roads are built precisely
because an area has been cleared and settled, rather than vice versa. And both the
land and the roads can be simultaneously influenced by a third set of factors, such as
soil quality or population density. Model makers have attempted to account for this
alternative by including some of those factors as separate independent variables, us-
ing road density, say, and analyzing only forest clearings that occur *after* roads are
built. These attempts have been only partially successful, but no policy intended to
influence deforestation can be considered comprehensive unless it provides clear
guidelines on investments in transportation infrastructure.

Property Regime and Tenure Security

In the absence of well-defined and secure property rights, forest clearing often be- comes a way to claim property rights to land (homesteading). Such strategic behav- ior has been reported by Anderson and Hill (1990), Mendelsohn (1994), and Angelsen (1999). Under these circumstances, there are at least three reasons why forests may be cleared *beyond* the point where the current net benefits are zero. First, even though profits may be negative in the first few years, technological progress, new roads, and so on will make cultivation profitable in the future, and farmers need to act now so that others do not claim the land before they do. Second, in many cases land prices may reflect not agricultural potential but rather speculation that the purchaser will profit from selling the land at some future date (Clark, Fulton, and Scott 1993). And finally, in situations where users compete for forest land, such as in conflicts between communities and government agencies, deforestation by one agent is costly to the other. Hence there may be incentives to clear the land oneself in order to squeeze out the competitor (Angelsen 1997).

Some empirical evidence suggests that where farmers can obtain property rights by clearing forests, land-titling projects can encourage them to clear larger areas (Kaimowitz 1996). Secure tenure encourages investment by making it less risky; if the investment involves clearing land in the forest, deforestation should increase as a result. Nevertheless, household- and regional-level regression models from Latin America show that deforestation is lower in areas with secure land tenure (Southgate, Sierra, and Brown 1991; Godoy and others 1996; Pichón 1997). Thus a conclusion is premature at this time.

Timber Prices

The literature on the effect of logging on deforestation is smaller, and the results less conclusive. The effect of higher timber prices remains particularly controver- Higher prices for timber are likely to promote deforestation by making logging more profitable (Capistrano 1990; Gullison and Losos 1993; von Amsberg 1994; Barbier and others 1995; Deacon 1995; Mæstad 1995). Higher timber values also increase the net benefits of clearing land (assuming the timber is sold) and encourage defores- tation (Southgate 1990; Deininger and Minten forthcoming).

Using a traditional supply-demand framework, trade restrictions, such as ex- port taxes and import bans, would reduce total demand for timber by lowering prices and production even if lower prices increased domestic demand. Other authors sug- gest, however, that in the medium term, low timber prices discourage efficient har- vesting and processing techniques, leading in turn to more logging (Barbier and others 1995). Low timber prices may also discourage efforts to prevent farmers from clearing logged areas (van Soest 1996).

he Underlying Causes of Deforestation

is harder to establish clear links between underlying causes and deforestation. acroeconomic variables influence decisions through complex paths, and many of e causal relations are indirect. Further, such studies typically require data that often do not exist or are of poor quality. Table 3 summarizes the major findings on the derlying causes of deforestation, with these reservations in mind.

opulation Pressures

eforestation rates may increase because the population is growing and needs more nd for food, fuelwood, timber, or other forest products. Growing populations also fect labor markets, as an abundant supply of labor pushes down wage rates. But

Table 3. *Major Results on Underlying Causes of Deforestation*

[illegible]	*Effect of increase in variable in model type* [illegible]		
	Theoretical [illegible]	*Simulation and empirical* [illegible]	*Comments*
[illegible]ulation	Increase	Increase	The empirical results suggest that population density is positively correlated with deforestation, but the evidence is weaker than often believed; regional population should be considered endogenous.
[illegible] level	Indeterminate	Increase	Higher income increases demand for agricultural and tropical products and access to markets but also increases off-farm employment.
[illegible]ome growth	Indeterminate	Mixed	Same as above
[illegible]ological progress [illegible] general equilibrium [illegible]	Reduce	Limited evidence	Should induce downward pressure on agricultural prices and upward pressure on wages and interest rates (unless the changes reduce labor and/or capital intensity)
[illegible] debt	Indeterminate	Mixed	Theory weak; empirical evidence weak and contradictory.
[illegible] liberalization [illegible] devaluation	Indeterminate	Increase[a]	Higher agricultural and timber prices increase clearing, but income declines may offset this in the short run; relative prices also matter.

[a] [illegible] may not be reliable
[illegible] Authors' analysis.

population growth may also induce technological progress and institutional changes that contribute to reduced pressures on forests.

Analytical models that consider the labor supply to be exogenous give quite different results from those that assume it to be highly elastic with respect to wages. In the former, deforestation rates tend to be much more sensitive to agricultural price changes, and agricultural intensification is more likely to diminish forest clearing (Angelsen 1999).

Several multicountry regression models show a positive correlation between population density and deforestation (such as Palo 1994; Rock 1996). Many of these results are spurious, however, because they rely on the FAO Forest Resource Assessments, which are themselves based on population data. As Rudel and Roper (1997: 54) note, "a variable which FAO used to construct the dependent variable is now being used to predict the value of that variable!" At the regional level, studies from Brazil (Andersen 1996; Pfaff 1997), Ecuador (Southgate, Sierra, and Brown 1991), Mexico (Barbier and Burgess 1996), the Philippines (Kummer and Sham 1994), and Thailand (Katila 1995; Cropper, Griffiths, and Mani 1997) also find a positive correlation between population density and deforestation. In the multicountry and regional studies, this correlation disappears when additional independent variables are added, implying that population may be acting as a proxy for some other factors in these models (Capistrano 1990; Deacon 1994; Harrison 1991).

The evidence on the relation between population growth and forest clearing is even weaker. Kimsey (1991) and Rock (1996) report that population growth increases deforestation. Burgess (1991) and Inman (1993) find that it reduces deforestation or has mixed effects, and Cropper and Griffiths (1994) and Palo (1994) that it has no effect.

Few models focus specifically on the relation between population and the demand for agricultural and forest products. Economic liberalization and globalization are likely to make this aspect less important at the national and regional levels because global demand is increasingly likely to determine prices and demand. New prospects for agricultural and forestry exports may lead to rapid deforestation in countries where small domestic markets previously limited deforestation.

At the local and regional levels, population is endogenous and is determined by infrastructure availability, soil quality, distance to markets, off-farm employment opportunities, and other factors. Several studies show that population growth in previously forested, low-population areas occurs in response to road construction, available high-quality soils, and growing demand for agricultural products (Harrison 1991; Southgate, Sierra, and Brown 1991; van Soest 1995; Andersen 1997). Government policies that affect migration (and hence population) at this level include road construction, colonization policies, agricultural subsidies and tax incentives, and gasoline prices. This implies that the latter factors, rather than population growth per se, are the causes of deforestation in these areas. People migrate to forested areas

ause clearing forest for agriculture is economically attractive, and so the size of population in those areas cannot be considered an independent variable in mod- of deforestation.

ome Level and Economic Growth

gher national income and economic growth can be expected to reduce the pres- e on forests by improving off-farm employment opportunities, but to increase it stimulating demand for agricultural and forest products and improving access to gin forests and markets. Countries with higher incomes may also demand that sts be protected rather than depleted. Forest depletion may contribute to eco- mic growth, implying a causal relation in the opposite direction.

Many studies of developing countries associate higher national per capita income h greater deforestation (Capistrano 1990; Burgess 1993; Krutilla, Hyde, and Barnes 95; Barbier and Burgess 1996; Mainardi 1996). Again, these models have signifi- n data and methodological weaknesses and should be regarded with caution. Evi- dence on the impact of income growth rates is even weaker. Because there is no ng short- or medium-term relation between economic growth rates and average apita national income, the fact that higher incomes are associated with more restation does not necessarily imply that higher growth rates will be.

he models are also not very clear about whether deforestation declines or is even sed beyond certain income levels as countries become richer, a possibility noted ie "forest transition" hypotheses (Mather 1992; Grainger 1995) and by the en- nmental Kuznets curve literature (for example, Stern, Common, and Barbier 6). Based on the dubious FAO data, several authors claim to have found an envi- nental Kuznets curve for deforestation: that is, at low levels of income, an in- e in income will accelerate the rate of deforestation, but higher income beyond tain level reduces deforestation. But the levels of per capita income they estimate t be reached before deforestation declines vary considerably (Panayotou 1993; pper and Griffiths 1994; Rock 1996). In addition, the driving forces behind a possible transition are still unclear. They could be economic forces (the attrac- of off-farm employment, a higher value placed on pristine forest by the public the government, or expanded state capacity to enforce forest protection). Even if relationship does exist, income levels in most tropical countries are well below level at which deforestation begins to decline.

rnal Debt, Trade, and Structural Adjustment

ne studies find a positive correlation between external indebtedness and defores- on (Burgess 1991; Kahn and McDonald 1994; Mainardi 1996; Kant and Redantz), while others find no clear connection (Capistrano 1990; Kimsey 1991; Inman

1993). The empirical studies are based on poor-quality data; the analytical models make very simplistic assumptions about government objectives and policy formation that limit their empirical relevance.

According to analytical models, policies to improve the terms of trade for agriculture tend to raise the prices received by farmers and hence increase deforestation (Jones and O'Neill 1994, 1995). Thus structural adjustment policies of this type may potentially increase pressure on forests, and policies such as overvalued exchange rates, industrial protectionism, and urban-biased spending may actually be good for forest conservation—although obviously not necessarily for other parts of the economy.

Market characteristics and general equilibrium effects can either strengthen or dampen these policy effects. Increases in agricultural and timber prices will generate more deforestation when labor supply is relatively elastic. If it is not, the initial effect of price increases will be dampened as rural wages rise in response to greater demand for labor.[4] Conversely, higher rural wages could potentially generate more demand for agricultural and forest products.

Structural adjustment and trade liberalization policies designed to increase the terms of trade in favor of agriculture may have short- or medium-term recessionary consequences that reduce urban food demand, which could lead to lower, rather than higher, agricultural prices and thus to less deforestation. But a recession might also lower urban employment, putting downward pressure on rural wages and consequently stimulating deforestation (Jones and O'Neill 1995).

Policies designed to increase agricultural and forest product exports are likely to affect deforestation more than policies that promote production for the domestic market (since the latter are more likely to exert downward pressure on prices). Similarly, pro-agricultural policies can be expected to have stronger deforestation effects in the contexts of globalized agricultural markets and trade liberalization.

The previous findings are supported by several analytical macroeconomic and computable general equilibrium models, which show that currency devaluation, trade liberalization, and agricultural subsidies increase deforestation (Cruz and Repetto 1992; Jones and O'Neill 1994, 1995; Wiebelt 1994; Barbier and Burgess 1996; Mwanawina and Sankhayan 1996). It should be remembered, however, that these models depend heavily on more or less arbitrary assumptions about price elasticities and use generally poor data. Moreover, all of them tend to look at the agricultural and forestry sectors at a very aggregated level. Changes in relative prices within these sectors may have a greater impact on deforestation than the overall sectoral terms of trade, and to date these models have shed little light on this subject.

These findings suggest the difficulties of evaluating the effects of macroeconomic policies; important effects are not included. For example, will increased public revenues give officials the leverage they need for better regulatory intervention? Or will affluence mean additional investments that increase forest clearance? One less-

any general claims about the relations among economic liberalization, structural istment, and deforestation are misleading. In particular, claims that structural istment programs will "generally contribute to both economic and environmen- gains" (Munasinghe and Cruz 1995) seem unjustified based on the evidence. If thing, the findings support the opposite claim because higher agricultural output timber prices lead to increased pressure on forests.

Indirect Effects of Technological Change

hnological inputs also have indirect (general equilibrium) effects on product, or, and factor markets. Technologies that increase aggregate supply and lower es should reduce pressures to clear additional forest land. In some cases this may n offset the initial effects of technology on deforestation, as is possible in the case naize production in the Philippines reported by Coxhead and Shively (1995). hnological changes that affect products with inelastic demands are more likely to uce deforestation. Labor-intensive technologies will raise rural wages and should mpen and even reverse the deforestation associated with the increased profit- ility of agriculture. In fact, the more labor-intensive the technology, the more rigid e labor supply, and the more prices of agricultural products respond to changes in sor costs, the greater will be the effect. Similarly capital intensive technologies might ve the same effect if farmers have limited access to capital.

Technologies such as irrigation that require substantial infrastructure and that nefit farmers with access to markets are particularly likely to reduce pressure on rests; they will tend to push down agricultural prices and bid up wages without reasing the profitability of frontier farming. At the empirical level, some studies nclude that technological progress leads to more deforestation (Katila 1995), while thers find the opposite (Panayotou and Sungsuwan 1994; Southgate 1994; Deininger nd Minten forthcoming).

Rethinking the Causes of Deforestation

his review raises serious questions concerning the conventional wisdom about the uses of deforestation, either by providing contrary evidence or by showing the eakness of the supporting evidence. In particular, the models raise significant doubts bout the following hypotheses:

- *The population thesis.* The models offer only weak support for the explanation hat population growth is a driving force of deforestation. The correlations are largely based on flawed data or incorrectly specified models. At the local and regional levels, population should be considered endogenous, particularly in the medium to long term.

- ***The poverty thesis.*** There is little empirical evidence on the link between deforestation and poverty. If forest clearing requires investment, rich people may in fact be in a better position to clear new forest land. Moreover, off-farm employment opportunities simultaneously affect both poverty and deforestation, and any apparent relation between poverty and deforestation may actually be reflecting the off-farm employment-deforestation connection. Poverty (and discount rates) should therefore be considered endogenous variables.
- ***The win-win thesis.*** The thesis advocated by the World Bank and others, [illegible] economic growth and the removal of market distortions are good for people and forests, finds limited support in this review. Economic liberalization [illegible] currency devaluations tend to yield higher agricultural and timber prices [illegible] in general, will promote deforestation. Moreover, higher incomes, within [illegible] relevant range of income found in developing countries, is likely to increase [illegible] pressure on forest resources.
- ***The making-the-forest-valuable thesis.*** Those who oppose boycotts of [illegible] timber and other timber market restrictions often claim that lower timber prices will discourage sound forest management. This review of the literature suggests that lower timber prices should both reduce logging activities and [illegible] agricultural encroachment stimulated by logging.
- ***The tenure security thesis.*** Land titles and more secure tenure have contradictory effects. Where forest clearing gives farmers a claim to the land [illegible] the security of such claims may lead to greater forest clearing. This [illegible] contradicts the conventional thesis of resource and environmental [illegible] that more secure property rights are good for the environment.
- ***The intensification thesis.*** How improvements in agricultural technology [illegible] forest clearing cannot be determined a priori, without information [illegible] the type of technology and the output and factor market elasticities. On [illegible] hand, intensification programs targeted at farmers living near the forest [illegible] make farming more profitable and may shift resources to forest clearing [illegible] attract new migrants, although this effect may be at least partially [illegible] by the resulting downward pressure on agricultural prices and upward [illegible] wages. On the other hand, new technologies for nonfrontier agriculture [illegible] reduce pressure on the agricultural frontier. Labor-intensive technological [illegible] are more likely to reduce pressure on forests than general yield-augmenting [illegible] ductivity increases and labor-saving technologies.

Although the evidence is not sufficient to reject all of these hypotheses, it [illegible] least raise significant doubts. It is time to rethink the causes of deforestation [illegible] redirect research to focus more on issues such as the impact of credit markets [illegible] nological change, poverty reduction, and land tenure.

Barbier, Edward B., N. Bockstael, J. C. Burgess, and I. Strand. 1995. "The Linkages between [illegible]ber Trade and Tropical Deforestation: Indonesia." *The World Economy* 18(3):411–42.

Binswanger, Hans P., Y. Mundlak, M. C. Yang, and A. Bowers. 1987. "On the Determinants [illegible] Cross-Country Aggregate Agricultural Supply." *Journal of Econometrics* 36:111–31.

Bluffstone, Randall A. 1995. "The Effect of Labor Market Performance on Deforestation in De[illegible]oping Countries under Open Access: An Example from Rural Nepal." *Journal of Environm[illegible] Economics and Management* 29:42–63.

Brown, Katrina, and David W. Pearce, eds. 1994. *The Causes of Tropical Deforestation [illegible] nomic and Statistical Analysis of Factors Giving Rise to the Loss of the Tropical Forests* [illegible] University College London Press.

Burgess, Joanne C. 1991. "Economic Analysis of Frontier Agricultural Expansion and [illegible] Deforestation." M.S. thesis. University College London, Environmental and Resource [illegible] ics. Processed.

———. 1993. "Timber Production, Timber Trade, and Tropical Deforestation." [illegible] 3):136–43.

Capistrano, A. D. 1990. "Macroeconomic Influences on Tropical Forest Depletion: A Cross [illegible] try Analysis." Ph.D. dissertation. University of Florida, Department of Food and Resource [illegible] nomics, Gainesville. Processed.

Chomitz, Kenneth M., and David A. Gray. 1996. "Roads, Lands, Markets, and Deforest[illegible] Spatial Model of Land Use in Belize." *The World Bank Economic Review* 10(3):487–[illegible]

Clark, J. S., M. Fulton, and J. T. Scott. 1993. "The Inconsistency of Land Values, Land [illegible] Capitalization Formulas." *American Journal of Agricultural Economics* 75:1[illegible]

Coxhead, Ian, and Sisira Jayasuriya. 1994. "Technical Change in Agriculture and Land De[illegible] tion in Developing Countries: A General Equilibrium Analysis." *Land Economics* [illegible]

Coxhead, Ian, and G. Shively. 1995. "Measuring the Environmental Impacts of Economic [illegible] The Case of Land Degradation in Philippine Agriculture." Agricultural Economics Staff Series 384. University of Wisconsin, Madison. Processed.

Cropper, Maureen, and Charles Griffiths. 1994. "The Interaction of Population Growth [illegible] ronmental Quality." *American Economic Review* 84:250–54.

Cropper, Maureen, Charles Griffiths, and Muthukumara Mani. 199[illegible] "Roads, Populat[illegible] sures, and Deforestation in Thailand, 1976–89." Policy Research Working Paper [illegible] Bank, Policy Research Department, Washington, D.C. Processed.

Cruz, Wilfrido, and Robert C. Repetto. 1992. *The Environmental Effects of Stabilization [illegible] tural Adjustment Programs: The Philippine Case.* Washington, D.C.: World Resources [illegible]

Dale, V. H., R. V. O'Neill, F. Southworth, and M. Pedlowski. 199[illegible] "Modeling [illegible] Management in the Brazilian Amazonian Settlement of Rondonia." *Conservation Biology* [illegible] 206.

Deacon, Robert T. 1994. "Deforestation and the Rule of Law in a Cross Section of Co[illegible]" *Land Economics* 70(4):414–30.

———. 1995. "Assessing the Relationship between Government Policy and Deforestat[illegible] nal of Environmental Economics and Management* 28:1–18.

Deininger, Klaus, and Bart Minten. Forthcoming. "Poverty, Policies, and Deforestation [illegible] of Mexico." *Economic Development and Cultural Change.*

Dvorak, Karen A. 1992. "Resource Management by West African Farmers and the [illegible] Shifting Cultivation." *American Journal of Agricultural Economics* 74:809–15.

Elnagheeb, A. H., and D. W. Bromley. 1994. "Extensification of Agriculture and Defore[illegible] Empirical Evidence from Sudan." *Agricultural Economics* 10:193–200.

FAO (Food and Agriculture Organization). 1981. *Tropical Forest Resources Assessment Project.* Rome.

———. 1992. *Forest Resources Assessment 1990, Tropical Countries.* Forestry Paper 112. Rome.

———. 1997. *State of the World's Forests, 1997.* Rome.

Foster, A. D., M. R. Rosenzweig, and J. R. Behrman. 1997. "Population Growth, Income Growth, and Deforestation: Management of Village Common Land in India." Brown University, Providence, and University of Pennsylvania, Philadelphia. Processed.

[illegible]kowski, James. 1997. "An Analytical Model of Deforestation Effects in Related Markets: The Case of Cocoa, Plantain, and Cocoyam Production in the ASB Cameroon Benchmark." International Institute for Tropical Agriculture, Yaoundé, Cameroon. Processed.

[illegible]dos, Ricardo, J. R. Franks, D. Wilkie, M. Alvarado, G. Gray-Molina, R. Roca, J. Escóbar, and M. Cardenas. 1996. "The Effects of Economic Development on Neotropical Deforestation: Household and Village Evidence from Amerindians in Bolivia." Harvard Institute for International Development Discussion Paper 540. Cambridge, Mass. Processed.

[illegible]dos, Ricardo, K. O'Neill, S. Groff, P. Kostishack, A. Cubas, J. Demmer, K. McSweeney, J. [illegible], D. Wilkie, N. Brokaw, and M. Martinez. 1997. "Household Determinants of Deforestation by Amerindians in Honduras." *World Development* 25:977–87.

[illegible] Alan. 1995. "The Forest Transition: An Alternative Approach." *Area* 27(3):242–51.

[illegible]son, R. L., and E. C. Losos. 1993. "The Role of Foreign Debt in Deforestation in Latin America." *Conservation Biology* 7(1):140–4[illegible].

[illegible]son, Susan. 1991. "Population Growth, Land Use, and Deforestation in Costa Rica, 1950–198[illegible]." *Interciencia* 16:83–93.

[illegible] Stein T. 1993. "Peasant Household Modeling: Farming Systems Evolution and Sustainability in Northern Zambia." *Agricultural Economics* 9:2[illegible]1–67.

———. 199[illegible]. "Adjustment Policies, Peasant Household Resource Allocation, and Deforestation in Northern Zambia: An Overview and Some Policy Conclusions." *Forum for Development Studies* [illegible]:[illegible]–34.

[illegible] Katherine. 1993. "Fueling Expansion in the Third World: Population, Development, Debt, and the Global Decline of Forests." *Society and Natural Resources* 6:[illegible]–39.

[illegible] D. W., and R. V. O'Neill. 1994. "Development Policies, Rural Land Use, and Tropical Deforestation." *Regional Science and Urban Economics* 24:753–7[illegible].

———. 1995. "Development Policies, Urban Unemployment, and Deforestation: The Role of Infrastructure and Tax Policy in a Two-Sector Model." *Journal of Regional Science* 35(1):135–53.

[illegible] D. W., V. H. Dale, T. J. Beauchamp, M. A. Pedlowski, and R. V. O'Neill. 1995. "Farming in [illegible]dônia." *Resource and Energy Economics* 17:155–88.

[illegible] R., and J. A. McDonald. 1994. "International Debt and Deforestation." In Brown and [illegible] eds.

[illegible]watz, David. 1996. *Livestock and Deforestation in Central America in the 1980s and 1990s: A Policy Perspective.* Bogor, Indonesia: Center for International Forestry Research.

[illegible]owitz, David, and Arild Angelsen. 1998. *Economic Models of Tropical Deforestation: A Review.* [illegible], Indonesia: Center for International Forestry Research.

[illegible] Shoshi, and Anke Redantz. 1997. "An Econometric Model of Tropical Deforestation." *Journal of Forestry Economics* 3(1):51–86.

[illegible] Marko. 1995. "Modeling Deforestation in Thailand." Paper presented at the IUFRO World Congress, Tampere, Finland. Processed.

[illegible] Mary. 1991. "An Analysis of Spatial and Temporal Variation in Tropical Deforestation." Ph.D. dissertation. University of Georgia, Department of Geography, Athens. Processed.

Krutilla, Kerry, William F. Hyde, and Douglas F. Barnes. 1995. "Urban Energy Consumption and Periurban Deforestation." *Forest Ecology and Management* 74(2):181–95.

Kummer, David M., and Chi Ho Sham. 1994. "The Causes of Tropical Deforestation: A Quantitative Analysis and Case Study from the Philippines." In Brown and Pearce, eds.

Lambin, Eric F. 1994. "Modeling Deforestation Processes: A Review." TREES (Tropical Ecosystem Environment Observation by Satellites). Research Report 1. European Commission Joint Research Centre/European Space Agency, Brussels. Processed.

Larson, Bruce A. 1991. "The Causes of Land Degradation along 'Spontaneously' Expanding Agricultural Frontiers in the Third World: Comment." *Land Economics* 67(2):260–66.

Liu, D. S., L. R. Iverson, and S. Brown. 1993. "Rates and Patterns of Deforestation in the Philippines: Applications of Geographic Information System Analysis." *Forest Ecology and Managem*[illegible] 57:1–16.

Ludeke, A. Kim, R. C. Maggio, and L. M. Reid. 1990. "An Analysis of Anthropogenic Deforestation Using Logistic Regression and GIS." *Journal of Environmental Management* 31:247–59.

Mæstad, Ottar. 1995. "Timber Trade Restrictions: The Effect on Tropical Deforestation." Discussion Paper 20/95. Norwegian School of Economics and Business Administration, Bergen. Processed.

Mainardi, Stefano. 1996. "An Econometric Analysis of Factors Affecting Tropical and Subtropical Deforestation." University of Natal, Department of Economics, South Africa. Processed.

Mather, Alexander S. 1992. "The Forest Transition." *Area* 24:36[illegible]9.

Mendelsohn, Robert. 1994. "Property Rights and Tropical Deforestation." *Oxford Economic Pa*[illegible] 46:750–56.

Mertens, B., and Eric F. Lambin. 1997. "Spatial Modeling of Deforestation in Southern Camero[illegible]" *Applied Geography* 17(2):1–19.

Monela, Gerald C. 1995. "Tropical Rainforest Deforestation, Biodiversity Benefits, and Sustain[illegible] Land Use: Analysis of Economic and Ecological Aspects Related to the Nguru Mountains, [illegible]zania." Ph.D. dissertation. Agricultural University of Norway, Department of Forest Sci[illegible] Aas. Processed.

Munasinghe, Mohan, and Wilfredo Cruz. 1995. *Economywide Policies and the Environment* [illegible] *from Experience.* Environment Paper 10. Washington, D.C.: World Bank.

Mwanawina, Inyambo, and Prem O. Sankhayan. 1996. "A Computable General Equil[illegible] Model to Measure the Effects of Structural Adjustment Programmes on Ecology and De[illegible]ment in Zambia." In *Structural Adjustment Policies and Environmental Degradation i*[illegible] *nia, Zambia, and Ethiopia. Preliminary Seminar Report.* Agricultural University of Nor[illegible] Aas.

Nelson, Gerald C., and Daniel Hellerstein. 1997. "Do Roads Cause Deforestation? Using Sa[illegible] Images in Econometric Analysis of Land Use." *American Journal of Agricultural Economi*[illegible] ruary):80–88.

Nghiep, Le Than. 1986. "The Land Rotation Farming System in Northern Brazil: Condition[illegible] Its Continuation and Transition to the Sedentary Cultivation System." Working Paper [illegible] 34. International Development Center of Japan, Tokyo. Processed.

Ozório de Almeida, Anna Luíza, and João S. Campari. 1995. *Sustainable Settlement in the Bra*[illegible] *Amazon.* New York: Oxford University Press.

Palo, Matti. 1994. "Population and Deforestation." In Brown and Pearce, eds.

Panayotou, Theodore. 1993. "Empirical Tests and Policy Analysis of Environmental Degrada[illegible] at Different Stages of Economic Development." Working Paper 238. International Labour [illegible]fice, Technology and Employment Programme, Geneva. Processed.

Panayotou, Theodore, and S. Sungsuwan. 1994. "An Econometric Analysis of the Causes of Tropical Deforestation: The Case of Northeast Thailand." In Brown and Pearce, eds.

Pearce, David W. 1996. "Global Environmental Value and the Tropical Forests: Demonstration and Capture." In W. L. Adamowicz and others, eds., *Forestry, Economics and the Environment.* Wallingford: CAB International.

[illegible]rsson, Annika, and Mohan Munasinghe. 1995. "Natural Resource Management and Economywide Policies in Costa Rica: A Computable General Equilibrium Modeling Approach." *The World Bank Economic Review* 9(2):259–85.

[illegible]aff, Alexander S. P. 1997. "What Drives Deforestation in the Brazilian Amazon? Evidence from Satellite and Socioeconomic Data." Policy Research Working Paper 1772. World Bank, Policy Research Department, Washington, D.C. Processed.

[illegible]chon, Francisco J. 1997. "Colonist Land Allocation Decisions, Land Use, and Deforestation in the Ecuadorian Amazon Frontier." *Economic Development and Cultural Change* 44(4):127–64.

[illegible]ck, Michael J. 1996. "The Stork, the Plow, Rural Social Structure, and Tropical Deforestation [illegible] Poor Countries?" *Ecological Economics* 18:113–31.

[illegible]sero-Bixby, Luis, and Albert Palloni. 1996. "Population and Deforestation in Costa Rica." Paper presented at the annual meeting of the Population Association of America, New Orleans. Processed.

[illegible] Ruerd, Gideon Kruseman, and Huib Hengsdijk. 1994. "Farm Household Modeling [illegible] Estimating the Effectiveness of Price Instruments on Sustainable Land Uses in the Atlantic Zone of Costa Rica." [illegible] Report 4. Wageningen Agricultural University, The Netherlands. Processed.

[illegible]der, Thomas K., with Bruce Horowitz. 1993. *Tropical Deforestation: Small Farmers and Local [illegible] in the Ecuadorian Amazon.* New York: Columbia University Press.

[illegible]der, Thomas K., and Jill Roper. 1996. "Regional Patterns and Historical Trends in Tropical Deforestation, 19[illegible]–1990: A Qualitative Comparative Analysis." *Ambio* 25(3, May):160–66.

[illegible]. 1997. "The Paths to Rain Forest Destruction: Cross-National Patterns of Tropical Deforestation, 1975–1990." *World Development* 25(1):53–65.

[illegible] S. A., and A. [illegible] Joyce. 1988. "Deforestation Rates and Trends in Costa Rica." *Biotropica* [illegible]

[illegible]gate, Douglas. 1990. "The Causes of Land Degradation along 'Spontaneously' Expanding [illegible]cultural Frontiers in the Third World." *Land Economics* 66(1):93–101.

[illegible]. 1994. "Tropical Deforestation and Agricultural Development in Latin America." In Brown and Pearce, eds.

[illegible]gate, Douglas, Rodrigo Sierra, and Laurence Brown. 1991. "The Causes of Tropical Deforestation in Ecuador: A Statistical Analysis." *World Development* 19(9):1145–51.

[illegible] Dan [illegible], M. S. Common, and E. B. Barbier. 1996. "Economic Growth and Environmental [illegible]gradation: The Environmental Kuznets Curve and Sustainable Development." *World Development* 24(7):1151–60.

[illegible] Rainer. 1995. "Conserving Tropical Rain Forests in Indonesia: A Quantitative Assessment [illegible] Alternative Policies." *Journal of Agricultural Economics* 46(2):187–200.

[illegible] Rainer, and Manfred Wiebelt. 1994. "Policies to Reduce Tropical Deforestation and Degradation: A Computable General Equilibrium Analysis for Cameroon." *Quarterly Journal of International Agriculture* 33(2):162–78.

[illegible] Lena. 1995. "Environmental Impact on Governmental Policies and External Shocks in [illegible]tswana: A Computable General Equilibrium Approach." In C. A. Perrings and others, eds., *[illegible]diversity Conservation.* Amsterdam: Kluwer Academic Publishers.

van Soest, Daan. 1995. "Tropical Rainforest Degradation in Cameroon." Tilburg University, Department of Economics, The Netherlands. Processed.

———. 1996. "The Impact of Trade Measures on Tropical Deforestation: A General Equilibrium Approach." Tilburg University, Department of Economics, The Netherlands. Processed.

von Amsberg, Joachim. 1994. "Economic Parameters of Deforestation." Policy Research Working Paper 1350. World Bank, Policy Research Department, Washington, D.C. Processed.

Wiebelt, Manfred. 1994. "Stopping Deforestation in the Amazon: Trade-off between Ecological and Economic Targets?" *Weltwirtschaftliches Archiv—Review of World Economics* 131(3):54?–68.

Formal and Informal Markets for Water: Institutions, Performance, and Constraints

K. William Easter • Mark W. Rosegrant • Ariel Dinar

Water markets—either formal or informal—can be an efficient method for reallocating scarce water supplies. At the same time certain constraints can raise the transaction costs of trading water. This paper reviews the conditions necessary to establish successful water markets, identifies potential problems, and offers mitigating strategies. It also uses examples of several informal and formal water markets already in operation to illustrate these problems and the solutions to them.

One response to growing demands for limited supplies of water is to reallocate available supplies through water marketing strategies. Although marketing is not a new concept, what is new is the growing recognition that the policy of developing new water supplies to meet future needs is no longer viable. Thus government officials are more open than ever before to new ideas for improving water management. In addition, recent estimates of the economic benefits from trading water within and among sectors illustrate the potential for relatively substantial gains from trade (Vaux and Howitt 1984; Colby 1990; Chang and Griffin 1992; Weinberg, Kling, and Wilen 1993; Rosegrant and Binswanger 1994; Hearne and Easter 1997; Thobani 1997; Hearne and Roe 1998). As a result countries such as Mexico have taken the plunge and established water use rights and water trading strategies as part of their new policy for managing water resources. Others, such as Peru and Pakistan, are considering or have considered implementing such programs (World Bank 1995, 1997).

This article proposes that countries facing water shortages under their current water pricing systems consider water marketing as a way to reallocate water resources. We illustrate the importance of understanding a country's institutional framework before embarking on a comprehensive overhaul of water policies and review the conditions required for effective water markets. Recent studies of formal and infor-

The World Bank Research Observer, vol. 14, no. 1 (February 1999), pp. 99–116

mal markets highlight the gains from the efficient allocation of water as well as th
constraints that raise the transaction costs of trading water. As we point out, wat
markets can provide the appropriate economic incentives to improve the efficienc
of water use and encourage the reallocation of water to higher-valued uses withou
encountering the traditional opposition of existing water users.

The Institutional Setting

If a country has little experience with private markets for allocating scarce goods an
services, water is unlikely to be one of the first goods exposed to private mark
forces. In contrast, in a country that is exploring new ways to use the private mark
to improve the allocation of publicly managed resources, scarce water may be a goo
candidate for market trading, depending on one's view of the requirements for mar
ket exchanges.

There are two distinctly different opinions about the institutional setting require
for efficient market exchanges: The neoclassical view posits that a legal system
required; a more pragmatic view emphasizes the importance of informal contrac
enforcement. Greif (1997:239–40), for example, observes:

> This neoclassical view that places the legal system at the center of contract enforcement in market economies has recently been criticized on the basis of evidence indicating that many contemporary exchange relations in the West and elsewhere are informal. The associated contract enforceability is not provided by the legal system but is based on reputation, general morality and personal trust within social networks. Empirical evidence indicates the importance of two distinct systems of informal contract enforcement: the individualistic system of informal contracts enforcement prevalent in the West, under which the reputation and morality of the individuals are key and the collectivist system of contract enforcement prevalent in most other societies, under which personal trust within the social network is critical

Cooter (1997) comes to a similar conclusion in reviewing the problem of con-
tracting and establishing a rule of law that is consistent with a country's social n

At least in the case of markets for irrigation water, both the formal neoclass
legal system and the informal system appear to be at work. The transfer of perma-
nent water rights seems to require the certainty that is provided by a legally bas
approach in which water rights are recorded and can be defended in court. Wat
transfers among districts are likely to change the amount of water that is returned t
streams and rivers, called the return flow, and a formal market may be required t
prevent losses of return flows to downstream users. If, however, the sales are tempo
rary, that is, for one season or less, and do not affect return flows, informal mark

ased on informal water rights can suffice. These informal sales will likely be among rmers in the same water district and in many cases among farmers served by the me canal. In addition, these sales are not likely to be anonymous, and enforcement f the contracts will not be provided by the legal system but rather will be based on eputation and personal trust. This suggests that to obtain more interdistrict or nterjurisdictional water trades, a country will have to develop legal water rights that an be verified and defended in court at a reasonable cost.

Informal water markets work fairly well for groundwater as long as recharge to treams is adequate and the market has a sufficient number of sellers (Palanisami and aster 1991; Shah 1993; Saleth 1998). The "tit for tat" game-theory enforcement trategy appears to work; if farmers do not pay, their future supplies will be cut, and f a seller does not deliver, the buyer can use another supplier. In one area of Gujarat, India, farmers have pipelines from three or four different suppliers coming to their fields (Shah 1993), and they can buy from the supplier who offers the best price and ervice. Shah found that "while the main beneficiaries of private investments in pipe-ines have been the buyers of water, early operators in the water business were moti-vated mainly by the desire to establish monopoly positions and to overcome topo-graphical constraints in supplying water to a large command" (pp. 61-62). In other words, although sellers are motivated by profits to sell water, the buyers may be the big beneficiaries.

If a country decides to establish a formal water market, the community of users eeds to support the concept as fair and beneficial. Thus the law must be written so hat the resulting allocation of rights is equitable. If the economic rents from water rading are concentrated in the hands of a few individuals or the negative effects on hird party users are large and unmitigated, the community is not likely to obey the aw. Cooter (1997:214) notes that "a modern economy needs effective laws to pro-mote cooperation among people. Yet, states enact many laws that few people obey. eople tend to disobey, or obey out of fear, laws that are not consistent with social orms and to obey laws that reflect social norms." In Pakistan, for example, farmers d to disobey the law against trading canal water. In contrast, the 1981 Chilean ater law that establishes private water use rights is widely obeyed because Chile not nly has a long record of private water development but also allocates water rights ased on past use (Hearne 1998b).

Conditions for Effective Water Markets

is effective formal markets are dependent on some basic institutional and organi-zational arrangements to overcome a number of potential market constraints and to prevent other associated problems from developing (see Garrido 1998a:tables 1 and 2). For example, tradable water rights or water use rights need to be separated from

land rights. In many cases, institutional arrangements will also be needed to deal with third-party effects that result from changes in return flows or declining economic activity in the region in which water sales originate. Adequate management and infrastructure will be needed for trades that are not in the immediate vicinity, such as trades between users on different canals. Countries must consider establishing mechanisms to prevent monopoly control over water and to avoid the overexploitation of groundwater. In both cases, however, these problems can be dealt with through the manner in which water rights are designed, quantified, allocated, monitored, and enforced (box 1). How to do this effectively for groundwater is a particularly vexing problem in many developing and industrial countries.

Box 1. *Constraints on Effective Trades in Unregulated Water Markets*

Potential problem[a]	*Frequency*[b]
Third-party effects from a decline in output and employment in the water exporting area (1, 2, 3, 6, 7)	I
Reduction or changes in return flows along with any changes in water quality (1, 3, 4, 6, 8, 9)	F
Added incentive to overdraft open-access groundwater stocks, damage the aquifer, and increase pumping costs (3, 4, 10, 12, 13, 14)	I
Increased costs of irrigation system for the remaining farmers (3, 10)	I
Drop in land values (3, 11)	I
Market power for large-scale buyers or sellers (1, 3, 15, 16)	I

Mitigating strategy

1. Require review and approval of transactions by public agency
2. Establish a fund to compensate third parties damaged in trading, financed by levies on water transactions
3. Limit trades to a percentage of water rights in a given area or community
4. Revise water rights downward
5. Grant water rights to those using return flows
6. Limit trading outside the river basin or sector to consumptive water use
7. Open litigation to nonholders of water rights
8. Tax or ban trading from upstream to downstream users
9. Set minimum instream flows to maintain aquatic ecosystems
10. Require buyers to pay a fee for the costs imposed on the irrigation system from which water is transferred
11. Require lending agency to clear permanent water sales
12. Adjudicate groundwater rights
13. Tax groundwater sales based on their scarcity value
14. Limit trading in areas with rapidly declining groundwater stocks
15. Provide for regulation of monopolies or expand supply options.
16. Aid small rights holders with legal fees and registration

a. Numbers in parentheses refer to appropriate mitigating strategies.
b I, infrequent. F, frequent

In many cases, governments will also need to reserve or buy some of the water ghts to preserve instream uses that have strong "public good" characteristics, such recreation, fish production, and the preservation of aquatic environments.[1] As owe (1998) points out, preservation has become a growing concern in the western nited States as the demand for recreation and environmental services has grown. alifornia has even reallocated water from irrigation to improve instream flows into e San Joaquin–Sacramento delta that will help preserve aquatic environments in e delta (Archibald and Renwick 1998; Howitt 1998).

Adequate information concerning water supplies and demands is a basic requireent for the efficient operation of markets. In Chile, for example, water user associans were essential in providing such information (Hearne 1998b). The central govnment generally has a comparative advantage in obtaining water supply data, though the users are better able to determine their own demand. User associations ith access to such data can be important conduits for information. Asymmetric formation is much less of a problem for water markets with strong user associans, but farmers may withhold information about their willingness to buy or sell ater in order to obtain more favorable prices.

Informal water markets may be a good alternative, particularly if water allocation the local level is a problem and the transaction costs of establishing formal markets e high (meaning the costs of enacting legislation, establishing institutional and ganizational arrangements for markets implementing trade arrangements, and onitoring and enforcing trades). Besides allowing water to be sold to the most roductive farmers, informal markets would give all farmers an incentive to use their ater more efficiently

ains from Water Markets and Organizational Constraints

he informal water markets that have evolved suggest that water users will buy and ll water even if such transactions are illegal or discouraged by governments (Renfro nd Sparling 1986; Shah 1993). Problems arise when governments are asked to help velop formal markets or allow informal markets to develop, particularly within vernment constructed irrigation projects. Because these are subsidized projects, any government officials maintain that the users should not be able to sell the ater at a profit and that poor farmers will be disadvantaged because they will lose cess to the water unless they pay higher prices. (Both Meinzen-Dick 1998 and mith 1998 dispute this claim, however.) Thus, even though water markets can hange the incentives for water users and improve allocation, organizational constraints may prevent their introduction (box 2).

One prerequisite for water marketing is some type of water or use right that can be bought or sold separately from ownership of the land. Such rights may be difficult to establish and are likely to be resisted by public water agencies that fear they will lose a

Box 2. *Constraints That Raise Transaction Costs in Unregulated Markets*

Potential problem[a]	*Frequency*[b]	*Mitigating strategy*
Incomplete or poorly defined water rights that are not separate from land (1, 2, 8, 10)	W	1. Register and secure water rights. 2. Use proxies of water use (land area and use) to define water rights.
Inadequate infrastructure, including conveyance and storage systems (3)	F	3. Invest in infrastructure. 4. Provide management training for irrigation agencies.
Inadequate water management, lack of water user associations (WUAs), or both (4, 5)	F	5. Provide water users incentives to organize WUAs. 6. Carry out education program explaining the benefits of water markets
Imperfect or asymmetric information about trading (5, 9, 10)	I	7. Tax unused water rights
The granting of more water rights than warranted by existing supplies (11, 12, 13)	I	8. Keep up-to-date single basinwide water rights registries. 9. Use public agencies or WUAs as clearing houses for trades
Sleeper or inactive water rights that might be sold and activated by water markets (7)	I	10. Aid small water rights holders with free legal protection and information
Inappropriate initial allocation of water rights that causes conflicts among water users (14)	F	11. Encourage spot and option markets for water 12. Revise all water rights downward
Reallocation by government agencies of water among and within sectors without compensating original users (1, 4, 16)	I	13. Define two types of rights with one senior to the other 14. Base allocation of water rights on past use and conduct an auction for any surplus
Opposition from farmers and environmental groups (6, 10, 15, also strategies 3, 6, 7, 8, 9, 10, 15 from box 1)	I	15. Use part of the water traded to enhance stream flows 16. Base water rights on shares of the water supply rather than on absolute volume

a. Numbers in parentheses refer to appropriate mitigating strategies.
b. I, infrequent; F, frequent; W, widespread.

great deal of power if they allocate water rights to users. Giving users water rig- means that system operators (the government officials) have the responsibility to de- liver water more or less when the users want it. In contrast, a government agency that retains the water rights can dictate the conditions under which farmers will receive water, including (in some cases) the necessary side payments. Making the water rights tradable creates an even greater dilemma for government agencies. To prevent loss of control over tradable water rights, the National Water Commission in Mexico and some of the water districts in the western United States limit trading among water districts. In Mexico a water user must obtain special government approval to sell water outside the district or jurisdiction, and any profits from the sale must accrue to the district and not to the seller. This regulation discourages interdistrict trading, but it also reduces the chance that trades will have deleterious third-party effects.

Even if the users set up and hire the management unit that allocates water, the
iit may have an incentive to use monopolistic power and discourage water trades
th other jurisdictions. If too much water is transferred out of the district, the
ulting shortage may reduce economic activity and make the irrigation system—or
rts of the system—difficult to operate effectively. Once most of the farmers along
anal have sold their water, the few remaining farmers who own water rights on the
nal may be difficult and expensive to serve. Thus, although water markets may
ange user incentives and encourage efficiency, the management of the system may
event trades or raise the transaction costs of interdistrict or interjurisdictional trades.
Cooter (1997) argues that an organization seeking to maximize the wealth of its
embers will behave monopolistically toward outsiders but efficiently toward insid-
s. By fixing prices, establishing jurisdictional territories, and withholding informa-
on from the public, the organization will seek to create monopoly power for its
embers in dealing with nonmembers. These motivations may affect water user
ganizations and other water entities that can use their infrastructure as a monopo-
stic tool to block trades to outsiders. Water user organizations can either say that
eir canals are used to capacity or charge such high transmission fees that the trades
come unprofitable.

The organizational problem appears to involve two important aspects: the resis-
nce to trading water between or among districts or jurisdictions, and the problem
establishing water rights and giving the users more control. Other problems that
se transaction costs include legal challenges by third parties claiming they might
damaged by a transfer, the lack of sufficient infrastructure to transfer water among
tential buyers, and the lack of an effective means for verifying and enforcing water
ghts. The question is whether taking action to reduce these transaction costs is in
e best interests of a country. If the answer is even a tentative "yes," then the second
estion is how to lower these costs.

Potential Problems and Mitigating Strategies

Clearly, one should not go to the expense of establishing water markets if water is not
arce or likely to become scarce in the foreseeable future. In Chile, for example,
here water markets have been encouraged, there is no trading in the southern re-
on because water is not scarce in that area. Water markets will be active only in
gions where water is scarce.

Formal water markets are less likely to develop if they are constrained by the high
st of institutional development and by the absence of the management and infra-
ructure needed to implement trades. In areas where it is costly to establish, allocate,
nd enforce water rights, markets will be slow to develop. This has been an impor-
nt factor in South Asia, where the presence of many small, fragmented farms makes
t difficult to establish individual water rights. In such cases, water rights may have to

be allocated to water user associations, as they are in in Mexico, or to villages. If water trading is really to expand, these countries may also need to take steps to improve both their infrastructure and their system management. Yet most of these latter improvements are needed even without the desire to establish markets.

In South Asia there has also been a general aversion to markets, especially for allocating basic resources such as water and land, because of the belief that markets will disadvantage low-income farmers. In practice, as Saleth (1998) and Meinzen-Dick (1998) show, even "poor" farmers benefit from market exchanges. Clearly, if there are too few sellers, buyers may be disadvantaged, but Saleth and Meinzen-Dick point out that social conditions in India and Pakistan, for example, tend to mitigate against such exploitation. Saleth finds little or no price discrimination in mature water markets or in situations where kinship and social relationships are strong.

One of the biggest problems in establishing water markets in irrigated areas concerns issues associated with changes in return flows and water-related economic activities. These issues are important in California, as discussed in Archibald and Renwick (1998) and Howitt (1998). When water is transferred into other sectors or out of a river basin, governments need to have mechanisms in place that require the traders to take these third-party effects into account. Archibald and Renwick note, however, that these mechanisms must be carefully designed or they will foreclose many socially beneficial water trading opportunities. Thus the government must ensure that the appropriate institutional and organizational arrangements are in place. Clearly, a strong legal system is an asset in establishing such arrangements. Other key assets are a history of private irrigation development, strong water user associations, and appropriate conveyance and storage systems. Hearne and Easter (1997) illustrate how investments in storage capacity and a flexible infrastructure lower the transaction costs of water trading in Chile.

In contrast, Hearne (1998a) shows how too much government involvement in reallocation decisions can prevent water markets from developing. In Mexico the National Water Commission has foreclosed any possibility of intersectoral market exchanges in several regions. A better strategy may be for the commission to act as a broker in facilitating water trades. That approach would mean a major change in the commission's function and is not likely to occur without strong political pressure from farmers and other interested parties.

Because the physical, institutional, organizational, and technical conditions that affect the performance of potential water markets vary so much, it is difficult to predict what will happen if water markets are introduced in a new area. The use of experimental markets, as suggested in Dinar and others (1998), may be a low-cost first step toward developing and evaluating alternative institutional arrangements. Although not widely used for water resources, experimental markets can capture some of the complexities involved in water markets.

perience with Water Markets

mal water markets specify the volume and share of water to be sold, either for a period of time or permanently. Informal markets usually involve the sale of unasured flows of surface water from a canal for a set period of time or of water nped from a well for a set number of hours. Although the units sold in informal rkets may not be metered, both the buyer and the seller have good information out the volume transferred. The key difference between the two markets is the way which the trade is enforced. If the users must self-enforce trades because no formal er rights exist that can be enforced through the legal or administrative system, the rket is informal. Formal water markets are usually found in North and South erica, whereas informal markets are prevalent in the irrigated areas of South Asia.

ormal Markets

oundwater markets are important for agricultural production and the distribun of water throughout the irrigated areas of South Asia. Saleth (1998) estimates t 20 percent of the owners of the 14.2 million pumpsets in India are likely to be olved in water trading. This means that water markets are providing water for out 6 million hectares, or 15 percent of the total area irrigated by groundwater. In istan a survey reported that 21 percent of well owners sold water (NESPAK 1991). In areas where dependable precipitation recharges the groundwater, the benefits buying and selling water from tubewells have increased farmers' income and production. The economic gains from groundwater markets reflect improved efficiency pump management, in reducing conveyance losses, and in farm-level water use. ese markets also increase access to irrigation, especially for smaller-scale farmers o do not own tubewells and cannot afford to invest in a well without a market for n water.

Meinzen Dick (1998), in one of the few studies estimating the economic returns m access to water markets, found that water markets increased the availability and iability of water supplies. Both yields and income rose for those who purchased ter particularly for those who also had access to canal water supplies. The highest lds and income, however, were still found among farmers who owned their own bewells and had access to canal water.

ARESTING OVERDRAFTS. Given that markets for the sale of groundwater draw an open-access resource (that is, one that is available for capture to anyone who access), it is not surprising that problems arise in areas with high demands and ited supplies. Farmers have an incentive to ignore the scarcity and buffer stock lue of the groundwater and pump until their cost of pumping equals the market

price of water (Ramasamy 1996).[2] Over time, the cost of pumping and the price of water rise as the groundwater level declines. For example, the overdraft (that is, water use in excess of recharge) in the Coimbatore District of India is almost 5,000 cubic meters a year. Ramasamy estimates that if the overpumping continues, it will mean a drop in total net returns to farmers of between $42 million and $69 million, a result of the increased costs of power necessitated by increased pumping and additional investment to deepen wells. Here is a case where informal markets may exacerbate the problem, and formal markets may not work any better unless water rights can be established and enforced in strict quantity terms. The problem is not the water markets but the lack of exclusive property rights for groundwater. To establish such rights, the number of wells and the amount of water to be pumped would have to be agreed on and restricted. Such restrictions are probably unrealistic without strong support in the irrigation community. If exclusive water rights can be established, however, the water market should reflect the scarcity value of water and help restrain overpumping.

Blomquist (1995) reports on one case where the demand for water is increasing and the community of water users has been able to stop the overdraft. In the dry Los Angeles metropolitan area in southern California, pumping is metered and priced so that users have an incentive to shift from local groundwater to more expensive but more plentiful imported water. Surface and imported water are stored and used to recharge the groundwater in the basin. One result has been a halt in saltwater intrusion from the ocean in the area's coastal groundwater basins. In some of these basins pumping rights have been defined, limited to the basin's average recharge, and made transferable to other users through sales.

A more typical case, reported by Shah (1993), is in coastal Gujarat, India. Here the overdraft of coastal aquifers has caused a decline in groundwater supplies in some areas and saltwater intrusion in others. Shah argues that any effective reduction in this overdraft is unlikely without good local leadership and the involvement of water user groups. He argues that "legal, quasi-legal, and organizational instruments of public policy will not, on their own, succeed in securing the compliance of farmers unless they are accompanied by measures aimed at affecting private returns to irrigation . . . or unless the structure of property rights on the water resource itself is drastically reformed (p. 147)." Similarly in Pakistan, Meinzen-Dick (1998: 218) doubts "whether government would have the institutional capacity to regulate sales and hundreds of thousands of private tubewells, and if it had such capacity, it is unclear what such direct intervention could achieve."

Yet in both India and Pakistan, any effect that water markets might have on the overdrafting of groundwater is much less than the effect of subsidized electricity. The zero or near-zero marginal cost of pumping means that farmers have an incentive to use water to the point where the marginal value of production is close to zero. This, of course, encourages farmers who can sell water to use their wells at close

full capacity. The low power rates not only create overdrafting problems but also waste electricity in countries without adequate power.

As noted above, water markets can actually help solve the overdrafting problem by increasing the incentives for efficient water use and making it possible to purchase water from areas where water is abundant. The ability to find another source of water, but at a higher marginal cost, can help promote community action for self-regulation and demand management. Shah (1993) cites a case in coastal Gujarat where self-regulation became possible when additional new supplies were piped into the area.

Overdrafting tends to be concentrated in coastal areas of India and Pakistan and in the hard rock areas of southern India. In many of the northern areas, pumping actually improves growing conditions by lowering the water table below the root zone (Shah 1993, Meinzen-Dick 1998). In cases where water tables are high or recharge rates are rapid, water markets are not likely to cause negative externalities except possibly temporarily if neighboring wells are too close or deep tubewells interfere with shallow wells. Where these externalities are small, personal trust and reputation may be enough to foster competitive informal water markets. This is particularly true where farmers own a number of separate plots that cannot be served by the same well. In such cases, most water sellers are also buyers because most farmers who own a well are able to irrigate only their large plots and must purchase water to irrigate other plots (Shah 1993; Meinzen-Dick 1998; Saleth 1998). In addition, their wells are likely to be underutilized unless they can sell water. Yet because of the costs of conveying water and the need for cooperation from neighboring farmers when water is to be conveyed any distance, high transaction costs can restrict trades in areas with only a few wells and prevent water markets from being competitive.

COUNTERING MONOPOLY PRICING This raises the other concern about water markets: the potential for monopoly pricing and discrimination. Groundwater markets are somewhat confined by the physical limits of the location and supply of groundwater. Still, pipelines can extend markets, and the investment costs of new wells would put a limit on monopoly power. An abusive monopolist who raises prices too high will find others investing in wells and undercutting the price. Shah (1993) cites a lack of balance between the numbers of buyers and sellers in areas with high-capacity wells, where one seller may serve as many as 70 or 80 buyers. He fails to say how many sellers the average individual buyer can access. Monopoly pricing may be avoided if the buyers can purchase water from three or four sellers— so long as the sellers do not collude.

The evidence on monopoly pricing is mixed. In a 1991–92 survey in Pakistan, Meinzen-Dick (1998) found that sellers were pricing water at little more than the cost of pumping. The two most common ways of charging for groundwater are a flat charge per hour of pumping (ranging from $0.57 to $3.27 an hour, depending on

the pump type, capacity, and location) and arrangements whereby the buyer supplies the diesel and motor oil for the pump and pays an additional fee of $0.16 to $0.24 an hour to the well owner to cover the wear and tear on the engine.[3] Sellers with diesel pumps were just recovering their own costs under either type of contract.

In contrast, Saleth (1998) suggests that in some areas of India, monopoly rents may be extractive. He cites as evidence the variation in water charges compared with pumping costs in different areas. For example, water charges are 1.3 to 2 times higher than operating costs in the Indo-Gangetic region but 2.5 to 3.5 times higher in the water-scarce hard rock regions of southern India. The difference in rates, however, might be explained in part by the difference in water scarcity and in the value of water in those two regions.

The degree of monopoly power may also be related to the terms of the transaction or contract for water. Not surprisingly, some of the contracts for water are quite similar to contracts for land. Water contracts include crop sharing, crop and input sharing, and labor arrangements. If the payment is cash-based, buyers have more freedom to take their business to another well owner anytime during the season. When the transaction is a contract in kind, especially one based on crop sharing or on crop and input sharing, the buyer is tied to the seller for at least one season if not longer. Similarly, if buyers contract to pay for the water with their labor, they may find it difficult to change suppliers until they have fulfilled the contract. Yet in most villages, informal markets do not appear to face extreme cases of monopoly rents. In fact, monopoly power that restrains trading in areas with serious problems of declining groundwater levels may help reduce overextraction. In contrast, when suppliers are taking advantage of their monopoly position and there are adequate groundwater supplies, the best strategy is to encourage (legalize) trading and increase competition through community and private well development (Palanisami and Easter 1991).

Thus informal water markets can improve water use and incomes in irrigated areas where water rights are not well defined or recorded. They also may be a good option if formal water markets are likely to produce third party challenges and result in excessively high transaction costs. Finally, informal markets would work well in traditional irrigation systems where the farmers manage the irrigation system and would be able to maintain a relatively modest level of transaction costs.

Formal Markets

In situations where informal markets can work well, it may not be necessary to incur the extra expense of establishing formal water markets. Formal markets will be required, however, to provide the certainty necessary for permanent water transfers and transactions between different sectors and jurisdictions. Because the need for permanent trades and interjurisdictional water exchanges is likely to become more important as nonagricultural demands for water grow, formal water markets are likely

become more common. The growing demand in water-scarce regions has been one of the driving forces behind the new interest in water markets. Several studies have illustrated the benefits that are possible from interjurisdictional trading in permanent water rights for short-term use.

In Texas 99 percent of the water traded has been transferred out of the agricultural sector in the Rio Grande Valley to nonagricultural users (Griffin 1998). Of the municipal water rights in the valley that existed in 1990, 45 percent had been purchased since 1970. Although water markets are not active in other areas in Texas, Griffin notes that the surface water law has evolved to the stage where trading will be more widespread in the future. In contrast, the groundwater law is just beginning to evolve.

ECONOMIC GAINS. In a study of the Guadalquivir Basin of southern Spain, Garrido (1998b) finds that the economic gains of trading within an individual water district or community may be relatively modest. In contrast, if permitted, trades among communities subject to different supply constraints and drought conditions could produce substantial gains. Garrido estimates the total welfare gain at no more than ...percent over the current water allocation for four communities where trades were only intracommunity. Intercommunity trading, however, could produce estimated economic gains in one of the older irrigation communities of almost 50 percent. Garrido also shows that both types of trades are very sensitive to the level of transaction costs. If those costs exceed 8 to 12 percent of the market price, trading and the gains from trading would be too small to justify the expense of establishing formal markets. Yet Garrido may underestimate the potential gains because he considers ...the crops traditionally grown in the region (cotton, wheat, corn, oilseed, and ...at beets) and excludes any transfers to nonirrigation uses. Evidence from Chile ...ed significant changes in cropping as a result of water trading (Hearne and Easter ...).

In contrast, Horbulyk and Lo (1998) found that most potential gains from introducing water markets in Canada's Alberta Province were likely to come from trades ...in a subbasin. They considered four subbasins and compared the current water ...ation situation with the allocation under four separate markets (one in each ...basin), as well as with a market encompassing the total basin. The four separate ...scenarios created 90 percent of the welfare gains that were obtained when unre...ted trading was allowed among the four subbasins. The urban sectors purchased ... of the water, except on the South Saskatchewan River, where the agricultural ... purchased additional water when market trading was allowed among the subbasins.

...G PATTERNS AND TRANSACTION COSTS. In their analysis of selected water ...kets in Chile, Hearne and Easter (1997) found trading both within and between ...tors. In the case of permanent transactions either within or between sectors, well-...blished water use rights that were recorded and recognized by the government

were critical in fostering trade. Several trades between farmers and the city of Serena were not consummated because of uncertainty regarding ownership of water rights. La Serena is a growing vacation destination located on the coast in a region some 400 kilometers north of Santiago. Rapid growth in demand has strai the city's water supply, particularly during the summer tourist season. The oper of water markets allowed the city to purchase water and delay development of water sources. Starting in 1992, the city's water company purchased enough wate increase its water supply by 28 percent. Additional purchases were made by stream households for domestic uses and by farmers.

Elsewhere in Chile, significant trading occurs in the Limari Valley for agricult purposes (the urban sector has adequate water). A survey of 37 farmers selling w and 19 farmers buying it reported transfers of rights to 9.2 million cubic meters gains from trade (measured as the difference between the value of water to the se before the sale and the value to the buyer after the sale) were, on average, $2 cubic meter ($3,045 an acre-foot), with a transaction cost of $0.069 a cubic m ($86 an acre-foot).[1] This sample was neither random nor complete, but the numl surveyed were large enough to show that the water market was very active and created significant gains from trade. The largest gains accrued to three grape prod ers who purchased 5.8 million cubic meters of water (63 percent of the total amo traded in the sample). In these active water markets, transaction costs were low did not seem to constrain trading. In other areas, such as the upper section of Maipo River that supplies the southwestern Santiago area and irrigates 100,000 h ares, the transaction costs are high, and trading is quite limited. The Maipo Riv divided into three sections for management and water trading. As a result rights are uncertain, and the lack of adjustable control structures raises transa costs and therefore limits trading (Hearne 1998b).

Similarly, Archibald and Renwick (1998) found that high transaction cost California limited a large number of potentially profitable trades. Two types of tra action costs were identified: administrative induced costs, which are explicit included in the price of water sold through the California Water Bank, and poli induced transaction costs, which stem from existing legal requirements design avoid injuring owners of water rights, damaging fish and wildlife, and creating ne tive third-party effects in exporting areas. Administrative-induced transaction c including the costs of locating buyers and sellers and negotiating quantities, tim and other terms of transfer, were $0.041 a cubic meter ($50 an acre-foot) in and $0.014 a cubic meter ($17.50 an acre-foot) in 1992 and 1994. Policy-indu transaction costs in the West range from $0.152 a cubic meter ($187 an acre-foot Colorado to $0.044 a cubic meter ($54 an acre-foot) in New Mexico, all states w less stringent state and federal transfer requirements than California. Policy indu transaction costs in this range would be as much as or more than the potential ga from trading in the California Water Bank (Archibald and Renwick 1998).

Because of high transaction costs in Colorado, Howe (1998) recommends shifting the administrative responsibility for water transfers from the water courts to the state Engineer's Office. He also recommends reserving or acquiring water for "public good" uses such as recreation, as well as making other changes to allow water to be marketed as freely in Colorado as it is in the neighboring states.

Colby (1998) suggests that the claims of Native Americans have the effect of imposing high transaction costs on water trading in many western rivers. She argues that even though markets do not work well with high transaction costs, when those costs are compared with the costs of litigated solutions, water markets look like a much better alternative.

Howitt (1998) reports that spot and options markets performed well during California's droughts in the 1990s. Even though these markets are a fairly recent phenomenon, he thinks they are a promising option for stabilizing available water supplies in California and other similar areas. Permanent shifts in demand, however, require a much more active formal market for water rights.

Conclusions

Contrary to the claims of many critics, water markets have worked and are likely to be a better mechanism for reallocating water than the alternative methods. There are both formal and informal water markets at work today. In addition, there are spot market sales, sales of permanent water rights, and leasing arrangements that are similar to those used for land including crop sharing and cash rents (Saleth 1998).

Where water is scarce and large amounts of the available water supplies were committed to particular uses a long time ago, the economic benefits from water markets are likely to be large. In contrast, if the allocation was made fairly recently, based on most highly valued uses of water, and new opportunities are not available, then gains will be much more modest, as is shown in the Spanish example developed ... (1998b).

For markets to be effective, transaction costs must be kept low. To keep these low, the appropriate institutional and organizational arrangements need to be ... as well as flexible infrastructure and management. As pointed out earlier, critical first step is to establish tradable water rights or water use rights separate ... land, as well as the mechanisms to deal with third-party effects.

... it is difficult to establish legally enforceable, permanent water rights, a "thick" ... market may provide almost the same security as ownership of permanent water ... In other words, the ability to buy the water needed at a reasonable price may ... enough security so that firms are willing to invest in enterprises that are ... ndent on this purchased water. A contingent water market can provide additional security so that firms can be assured of a given volume of water at a set price.

With only a spot market and no contingent markets, firms may be subject to wide fluctuations in prices.

For those users needing certain supplies, spot water markets are probably cheaper alternatives than having to buy enough senior water rights so that one is guaranteed adequate supplies even in the worst drought. (Owners of senior water rights have the right to whatever water is available, before the more junior water rights owners.) In Pakistan, for example, the markets for groundwater have greatly improved the security of water supply, particularly in government irrigation projects. This security has allowed increased investment and increased production. It will be important to see if spot and contingent markets have similar effects on the productivity of water.

Finally, the evidence indicates that appropriately designed water markets, supported by sound institutions, are an effective mechanism for reallocating scarce water among sectors. Carefully designed water markets make it possible to meet the growing urban and industrial water demands without derailing growth in crop production. Market transfers among sectors may make it possible to significantly scale back investments in new water supply projects. Government inaction, ineffective institutions for water management, and high transaction costs, however, are likely to prevent water markets from reaching their full potential for reallocating scarce water resources.

Notes

K. William Easter is a professor of applied economics at the University of Minnesota, St. Paul; M. W. Rosegrant is a research fellow at the International Food Policy Research Institute, Washington, D.C.; and Ariel Dinar is a principal economist in the Rural Development Department of the World Bank.

1. "Public goods" are goods that consumers cannot be excluded from using and that are consumed during use but continue to provide the same benefits to other consumers (World Bank 1993).
2. Scarcity value is the opportunity cost of water. It is the present value of the sacrifices imposed on the future by using the resource today. Buffer stock value is the value of groundwater in stabilizing water supplies when the supply of surface water is uncertain (Tsur 1990).
3. The 1995 exchange rate of 24.5:1 was used to convert Pakistani rupees to U.S. dollars.
4. An acre-foot equals 1,233 cubic meters.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Archibald, Sandra O., and Mary E. Renwick. 1998. "Expected Transaction Costs and Incentives for Water Market Development." In Easter, Rosegrant, and Dinar.

NESPAK (National Engineering Services of Pakistan). 1991. *Contributions of Private Tubewells in t Development of Water Potential.* Field Report. Islamabad: Government of Pakistan, Ministry Planning and Development. Processed.

Palanisami, K., and K. William Easter. 1991. "Hydro-Economic Interaction between Tank Stor and Groundwater Recharge." *Indian Journal of Agricultural Economics* 46(2):174–79.

Ramasamy, Chellappan. 1996. "Intertemporal Allocation of Groundwater Resources with Stoch tic Surface Supplies: The Case of Southern India." Ph.D. thesis. University of Minneapo Department of Applied Economics, St. Paul. Processed.

Renfro, Raymond Z. H., and Edward W. Sparling. 1986. "Private Tubewells and Canal W Trading of Pakistan Punjab Watercourses." In K. William Easter, ed., *Irrigation Investment nology, and Management Strategies for Development.* Boulder, Colo.: Westview Press

Rosegrant, Mark W., and Hans Binswanger. 1994. "Markets in Tradable Water Rights: Pot for Efficiency Gains in Developing Country Irrigation." *World Development* 22:1613–25

Saleth, Maria R. 1998. "Water Markets in India: Economic and Institutional Aspects." In Rosegrant, and Dinar.

Shah, Tushaar. 1993. *Groundwater Markets and Irrigation Development: Political Economy and tical Policy.* Bombay: Oxford University Press.

Thobani, Mateen. 1997. "Formal Water Markets: Why, When, and How to Introduce Trad Water Rights." *The World Bank Research Observer* 12(2):161–79.

Tsur, Yacov. 1990. "The Stabilization Role of Groundwater When Surface Water Supplie Uncertain: The Implications for Groundwater Development." *Water Resources Res* 811–18.

Vaux, Henry, and Richard Howitt. 1984. "Managing Water Scarcity: An Evaluation of Inte gional Transfers." *Water Resources Research* (20):785–92.

Weinberg, Maria, Catherine Kling, and James Wilen. 1993. "Water Markets and Water Qu *American Journal of Agricultural Economics* 75(2):278–91.

World Bank. 1993. *Water Resources Management.* A World Bank Policy Paper. Washington, D

———. 1995. "Peru: A User-Based Approach to Water Management and Irrigation Develop Report 13642-PE. Washington, D.C. Processed.

———. 1997. "Pakistan National Drainage Program (NDP) Project." Staff Appraisal Report PAK. Rural Development Sector Management Unit, South Asia Region. Washington, D.C. cessed.

Judicial Reform and Economic velopment: A Survey of the Issues

Richard E. Messick

dging the importance of sound judicial systems to good governance and eco- wth, the World Bank and several other donor organizations have funded judi- n projects in more than two dozen developing countries and transition econo- ng the past few years. Yet little is known about the actual effect of judicial economic performance or even about what elements constitute a sound reform his article surveys a wide range of current studies on judicial reform and finds ising results

gnition that good governance is essential for economic growth has sparked interest in projects to reform judicial systems. Since 1994 the World Bank, American Development Bank (IDB), and the Asian Development Bank er approved or initiated more than $500 million in loans for judicial reform n 26 countries (Armstrong 1998). The U.S. Agency for International De- nt (USAID) has spent close to $200 million on similar projects in the past AO 1993), and other government and private groups are also funding pro- modernize the judicial branch of government (ACC 1995; Blair and Hansen ezger 1997). Today, the majority of developing countries and former so- tes are receiving assistance of some kind to help reform courts, prosecutors' nd the other institutions that together constitute the judicial system.

gh few now question the importance of judicial reform for development, nown about the impact of the judicial system on economic performance. ere any agreement on what makes for a successful judicial reform project. gue that reform cannot be achieved without a societywide consensus, while ntend that the reform project can help create this consensus. Fears are also ressed that judicial reform programs will repeat the mistakes of the law and ent movement, an earlier, American-sponsored initiative that unsuccess- ght to export the U.S. legal system wholesale to the developing world

nk Research Observer, vol. 14, no. 1 (February 1999), pp. 117–36
International Bank for Reconstruction and Development / THE WORLD BANK

The design and implementation of judicial reform projects are complicated by lack of knowledge about the relationship between formal enforcement of the law through the courts and traditional, extralegal—or informal—means of enforcemen Coincident with the growing emphasis on judicial reform, a body of research h emerged showing that the formal legal system is just one way of ensuring complian with society's laws. A variety of studies, in settings as diverse as medieval Europe an contemporary Asia, show that informal mechanisms based on incentives provide by repeat dealings can ensure the performance of contracts that no court has th power to enforce. One early, and surprising, finding of this research is that in som instances the sudden introduction of a formal mechanism to resolve legal disput can disrupt informal mechanisms without providing offsetting gains.

Rationales for Judicial Reform

Judicial reform is part of a larger effort to make the legal systems in developin countries and transition economies more market friendly. This broader legal refor movement encompasses everything from writing or revising commercial codes, bank ruptcy statutes, and company laws through overhauling regulatory agencies and teach ing justice ministry officials how to draft legislation that fosters private investme Although the line between judicial and legal reform blurs at the margin, the cor a judicial reform program typically consists of measures to strengthen the judic branch of government and such related entities as the public prosecutor and publ defender offices, bar associations, and law schools (Blair and Hansen 1994; Dako 1996; IDB 1994; Shihata 1995).[1] These measures aim to:

- *Make the judicial branch independent.* Included here are changes in the ways which judges are selected, evaluated, and disciplined to ensure that decisions insulated from improper influences. In some cases, the budget for the judic branch or the authority to administer the funds allocated for the judicial func is transferred from the ministry of justice or other executive branch agency to judges themselves. Independence can also encompass giving judges the powe declare acts of the executive and legislative branches of government in violat of the country's constitution or some other higher law.
- *Speed the processing of cases.* Providing management training, computers, and other resources to judges and court personnel reduces case backlogs and acce erates the disposition of new disputes. Revising the procedures for filing and resolving lawsuits helps to weed out procedures that invite delay and [illegible] costs.
- *Increase access to dispute resolution mechanisms.* The creation of mediation and conciliation services and other alternatives to resolving disputes in the courts

;es court costs, as does introduction of small claims courts or justices of the · and the establishment of legal aid societies. Actions may also include ferring responsibility for noncontentious matters, such as name changes, ırobate of uncontested wills, and the registration of property, to admin-ive agencies so that the courts have more time for disputed cases.

ssionalize the bench and bar. In-service training for judges, lawyers, and legal professionals entails programs to establish codes of ethics and disci-ry procedures. Increasing the number of law schools, ensuring that these ›ls have adequate resources, and modifying the curriculum to reflect the ınds of a market economy are also a part of this element.

ïunds judicial reform as part of its larger effort to strengthen newly emerg-:racies around the globe (Blair and Hansen 1994; GAO 1993; Walker 1995). ty's projects originated in the early 1980s to assist the then fragile demo-ernment in El Salvador to bring individuals accused of human rights abuses A small program was initiated to train judges and law enforcement person-estigative techniques. The program was subsequently expanded and then first to the rest of Latin America and later to the nations of Central and urope and the newly independent states of the former Soviet Union.

B finances judicial reform projects for a combination of political and eco-isons (IDB 1995). On the one hand it sees judicial reform as an indispens-ent in consolidating democratic institutions in Latin America by protect-human rights and promoting harmonious social relations. At the same ·cognizes that a well-functioning judicial system is important to the devel-›f a successful market economy. Judicial reform is part of the IDB's recent to help borrower countries modernize the machinery of government. Since ıtive was launched, the IDB has either approved or taken under consider-separate reform projects (Armstrong 1998).

il reform projects sponsored by the World Bank aim solely at enhancing a ·conomic performance. The Bank is enjoined by its Articles of Agreement ·rfering in the political affairs of its members, a prohibition it interprets as ıg it from supporting judicial reform unless the project "is relevant to the economic development and to the success of the Bank's lending strategy ›untry" (Shihata 1995:170). In practice this means that it does not provide to reform criminal codes, train police or criminal court judges, or manage ıtutions (World Bank 1995). This focus on the economic consequences of :form has led to complaints that such projects are ineffectual. Becker (1997) at the emphasis on narrow technical issues comes at the expense of more ıt, but arguably political, questions. (For a recent critique of a World Bank-d project in Venezuela by two human rights groups, see Lawyers Commit-.)

Judicial Reform and Economic Development

Whatever the rationale for judicial reform, it is widely believed that reform will significantly improve economic performance (Sherwood 1995). One hypothesis focuses broadly on the importance of the judicial system in enforcing property rights, checking abuses of government power, and otherwise upholding the rule of law. A second, narrower one casts the relationship solely in terms of the judiciary's effect in enabling exchanges between private parties. Although neither hypothesis has been subjected to a rigorous empirical test, there is some indirect evidence, albeit tentative and inconclusive, supporting both.

The narrow hypothesis originates with the 16th century English philosopher Thomas Hobbes, who argued that without a judicial system, traders would be reluctant to enter into wealth-enhancing exchanges for fear that the bargain would not be honored. In Hobbes's words, when two parties enter into a contract, "he that performeth first has no assurance the other will perform after because the bonds of words are too weak to bridle men's ambitions, avarice, anger, and other passions without the fear of some coercive power" ([1651] 1962:8).

Twentieth century development economists have revived Hobbes's thesis. North (1990:54) asserts that the absence of low-cost means of enforcing contracts is the most important source of both historical stagnation and contemporary underdevelopment in the Third World." In Williamson's (1995) view, a "high-performance economy" is one that is characterized by a significant number of long-term contracts—just the type of business relationship that is unlikely to thrive in the absence of a well-functioning judicial system. When the judiciary is unable to enforce contract obligations, a disproportionately large number of transactions takes place in the spot market, where there is less opportunity for breaching contracts. Or alternatively, firms circumvent the judicial system altogether by vertical and conglomerate integration, turning arms-length transactions into intrafirm ones. In either case, argues Williamson, the results are higher transaction costs and a "low performance economy."

Survey evidence from Ghana supports Williamson's argument that the absence of a judicial system raises transaction costs—but not in the way he posited. As reported in Fafchamps (1996), businesses in Ghana rely upon a network of traders to serve as go-betweens. Rather than solicit a supply of lumber, say, from an unknown company directly, a firm will enlist a trader that it knows and that knows the lumber company. The personal relationships provide the buyer and seller with some assurance that the lumber will be delivered and payment received, but at a price: the reliance on intermediaries raises the costs of doing business.

Constructing a direct empirical test of Hobbes's hypothesis is a formidable task. As Sherwood, Shepherd, and de Souza (1994) note, it would require determining what transactions are not taking place and then quantifying the resulting losses. Clague

and others (1995) proposed an indirect test instead, based on the assumption that the greater the percentage of money held in bank accounts and other financial assets, the more confidence citizens have in the judiciary and other institutions required to enforce bargains. Conversely, they reasoned, when a large percentage of the money supply is held outside banks and other financial institutions, the greater the likelihood that a substantial number of exchanges are consummated in simultaneous, spot-market transactions.

Cross-country regressions for a large sample of industrial and developing countries using a measure of the stock of money held in the financial system yielded the predicted results. The greater the percentage of the economy's money in the system, the higher the level of investment and, to some extent, growth. But as Castelar Pinheiro (1998) observes, these results are problematic. A large portion of the money held in financial institutions consists of currency and other liquid assets available for immediate withdrawal. Hence, it cannot readily be assumed that a high ratio of funds held inside the system necessarily means that fewer spot or simultaneous transactions are taking place.

Surveys of Latin American entrepreneurs provide somewhat more, if also indirect, support for Hobbes's thesis. In Peru almost a third of those responding to a World Bank poll said they would not switch from a trusted supplier to a new one—even if a lower price were offered—for fear the new supplier could not be held to the bargain (Dakolias 1996). A similar survey in Ecuador found that businesses were hesitant to invest because of the uncertainty of and potential lack of timeliness in enforcing contract rights. In-depth interviews of Brazilian entrepreneurs suggest that domestic investment would increase 10 percent if the Brazilian judiciary were on a par with those in the advanced market economies (Castelar Pinheiro 1998).

A second—far broader—hypothesis posits a more complex relationship between judicial reform and economic development. This view holds that economic development depends on a legal system in which not only are contracts between private parties enforced, but the property rights of foreign and domestic investors are respected and the executive and legislative branches of government operate within a known framework of rules (Dakolias 1996; Shihata 1995; World Bank 1992, 1994, 1997). This way of defining the rule of law assigns a prominent place to the judicial system: "[T]he judiciary [is] in a unique position to support sustainable development by holding the other two branches accountable for their decisions and underpinning the credibility of the overall business and political environment" (World Bank 1997:100).

The argument that the rule of law fosters economic development has been made many times. The 15th century jurist John Fortescue ([1471?] 1979) asserted that medieval England's prosperity was traceable to the quality of English legal institutions. Almost 300 years later Adam Smith ([1755] 1980:322) observed that "a tolerable administration of justice," along with peace and low taxes, was all that was

necessary to "carry a state to the highest degree of opulence." Max Weber, the 19th century German sociologist, was the first to look carefully at the relationship among the rule of law, a well-functioning judiciary, and economic development (Trubek 1972), but according to Hayek (1960), credit for recognizing the judiciary's importance in enforcing the rule of law belongs to the writers of the American Constitution and the German philosophers who elaborated the concept of the *Rechtsstaat*.[2] The former showed why judicial review of legislative actions was crucial, while the latter demonstrated the importance of subjecting the actions of the executive and its administrative agencies to judicial scrutiny.

Weber's comparative analysis of the role of law in China and the West was perhaps the first systematic effort to develop empirical support for the claim of a relationship (Bendix 1960). Most recently, in a survey of 3,600 firms in 69 countries, more than 70 percent of the respondents said that an unpredictable judiciary was a major problem "in their business operations" (World Bank 1997:36). The report also found that the overall level of confidence in the institutions of the government, including the judicial system, correlated with the level of investment and measures of economic performance.

But rigorous econometric methods for verifying the rule-of-law hypothesis re the role played by the judicial system are still in their infancy. Castelar Pinheiro (1998) reviews three recent efforts using cross-country regression analysis (Borner, Brunetti, and Weder (1995), Knack and Keefer (1995), and Mauro (1995). Each uses a proxy for judicial system performance, such as entrepreneurs' perception of the political risk involved in conducting business in a given country, to explore the correlation between a better judicial system and higher rates of investment, growth, and other indicators of economic performance. All three studies report a relationship between the proxies selected and different indicators of economic development, but as Castelar Pinheiro notes, each suffers from several methodological problems that make the results suggestive at best. The proxies for judicial system performance are often questionable, and there are problems with the endogeneity of the independent variables. These studies also do not rule out competing explanations such as increases in trade and investment or even the effects of other institutional reforms such as the introduction of an independent central bank.

Nor do cross-country regressions settle the question of the direction of causation. It may be that higher levels of development permit the state to spend more on the judicial system (Posner 1998). Or as Pistor (1995) observed in a review of judicial and economic reform in the transition economies, the same factors that contribute to economic reform and development may also be responsible for improvements in the judiciary. Both may be a result of preexisting attitudes and beliefs in society at large, or what has recently been termed "social capital" (Ellickson 1997; Solow 1995; World Bank 1997).

Hirschman's (1994) suggestion about the relationship between political and eco-
omic progress may apply equally to the relationship between judicial reform and
levelopment. He argues that political and economic progress are not tied together in
ny straightforward functional way. Rather, given the historical record, the relation-
hip is probably better modeled as a series of on-and-off connections, or of couplings
nd decouplings. At some stages in the development process, the two may be inter-
lependent, while at other stages they may be autonomous. There is no reason not to
elieve that a similar dynamic may be at work in the interplay between the evolution
of the judiciary and economic growth, and the legal transplant school of comparative
aw has marshaled an enormous body of evidence showing that substantive law de-
elops independent of economic and social variables (Ewald 1995).

In sum, while history and comparative analysis support the view that a better
udicial system fosters economic growth, there is, as Weder (1995) observes, no clear,
mpirical evidence showing the economic impact of a weak judicial system. The
most that can be said at the moment is that the weight of opinion and evidence
uggests the existence of some type of relationship.

The Prerequisites of Successful Judicial Reform

Judicial reform can threaten those with a stake in the status quo. As both Eyzaguirre
1996) and Blair and Hansen (1994) note, inefficiencies in court procedures and
management often provide opportunities for rent-seeking by attorneys, judges, and
judicial support personnel. In Argentina, for example, the judicial clerks have pro-
tested a proposal by Fundacion de Investigaciones Económicas Latinoamericanas
(FIEL 1996) that they work more than the current 132 days a year. (The increase
would raise the work year to at least 163 days, the average for executive branch
personnel, if not to the 231 average for Argentine private sector employees.) The
support staff is also challenging a recommendation to curb their power over case
management and courtroom scheduling.

Reform may also engender opposition from the nation's organized bar. In Uru-
guay lawyers objected to the introduction of procedures that would speed up civil
and criminal trials, fearing that speedier trials would mean less work for them (Vargas
1996). Reform can threaten lawyers' incomes in other ways as well. The practice of
law is almost invariably a state-sanctioned guild or cartel, but as Posner (1995) ex-
plains, unlike an oil or steel cartel, "legal services" are difficult to define. The state
must therefore specify what tasks are for lawyers and what tasks can be performed
just as well by nonlawyers. In Peru, for example, attorneys and public notaries vigor-
ously opposed measures to cut the costs of registering land belonging to the urban
poor because the measures would allow engineers, architects, and other professionals

to provide services that had once been the exclusive preserve of the legal profession (World Bank 1997).

Given the opposition that judicial reform is certain to generate, one view holds that no program should be undertaken absent a broad consensus in the country on the need for significant change. Dakolias (1996) recommends extensive consultation with committees representing judges, members of the bar, and other affected groups during the preparation and implementation of the project. Shihata (1995) adds that this consensus must include a long-term commitment on the part of the government to provide the resources required for an effective judiciary.

Blair and Hansen (1994) reached a similar conclusion in an evaluation for USAID of judicial reform projects in Argentina, Colombia, Honduras, the Philippines, Sri Lanka, and Uruguay. Absent a high level of support from the ministry of justice, senior executive branch officials, legislators, and judges, the authors argue that judicial reform is unlikely to succeed. When such support is lacking, they recommend that both public and private donors forgo judicial reform altogether. Instead they suggest that donors concentrate on building a consensus for reform by opening a dialogue with the government and by encouraging bar associations, business groups, and other nongovernmental organizations to campaign publicly for reform.

In the six cases examined, Blair and Hansen found that training judges, improving management systems, and supplying computers and other resources to the judiciary had little impact in countries where a consensus for judicial reform was lacking. The lesson they draw is that these traditional components of judicial reform, often termed "institutional strengthening," should not be initiated until more basic reforms have been achieved. Legal changes permitting the use of alternative dispute resolution mechanisms, creating or broadening legal aid programs, and ensuring that judges are appointed on the basis of merit should come first. Only after structural reforms and access-enhancing measures are in place do they support institutional strengthening.

Several other evaluations of judicial reform in Latin America appeared to confirm Blair and Hansen's findings. Even before their results were published, the U.S. General Accounting Office (GAO 1993), in an analysis of USAID-sponsored programs in Central American and Colombia, concluded that providing computers, [illegible] and other technical assistance to the judiciary in countries where a strong commitment to reform was lacking had not been productive. Buscaglia, Dakolias, and Ratliff (1995) report that Latin American judges often questioned the value of training in the absence of more fundamental reforms in the judicial system. In many cases, once a judge had been trained, he or she quickly left the bench for a more lucrative position in the private bar.

In a study of judicial reform projects in 15 Latin American countries, Martínez Neira (1996) found there had been too much emphasis on increasing the number of judges, courts, buildings, and computers at the expense of more fundamental changes

the legal system. He contends that this imbalance resulted from a lack of consen-
, on the scope of reform. Without such a consensus, judges, clerks, attorneys, and
ner actors in the legal system are free to pursue their own agendas. Pérez Perdomo
993) makes a similar point, arguing that too many Latin American reform pro-
ims reflect only the needs and perspectives of judges and others with an institu-
nal role in the judicial system.

Not everyone agrees with these criticisms, however. In a response to the General
counting Office's critique, USAID (1993) contended that such programs can be
en as the vehicles for developing a consensus. The collaboration between outside
perts and judges and others within the country and the ensuing public discussion
n help to generate the necessary political commitment, the agency maintained.

Hammergren (1998), who subscribes to this view as well, also takes issue with
ai and Hansen's recommendation that institutional strengthening should always
llow structural reforms and measures to increase access to the judicial system. She
serves that institutional strengthening can pave the way for broader reforms that,
attempted first, may engender such strong opposition that reform will be stalled.
e argues that institutional strengthening measures do not have to end up simply
rving the needs of judges, lawyers, and others with a stake in the status quo. Such
easures can include ways of making the judicial system more accountable to the
blic and, as MacLean (1996) has stressed, give it a public service orientation.
ammergren also cautions that nongovernmental organizations have their own in-
rests that may be at odds with a broader reform agenda. Business groups may, for
ample, be interested solely in the creation of commercial courts or steps that re-
ce legal fees.

he Law and Development Movement

e the legal technical assistance programs sponsored by the World Bank ignoring
e lessons learned in earlier attempts to foster development through law (McAuslan
" Thome 1997)? In the 1960s USAID, the Ford Foundation, and other private
ican donors underwrote an ambitious effort to reform the judicial systems and
merce laws of countries in Asia, Africa, and Latin America. This "law and devel-
nt" movement engaged professors from Harvard, Yale, Stanford, Wisconsin,
other leading law schools and within a few years had generated hundreds of
ts on the contribution of law reform to economic development (Merryman
). Yet after little more than a decade, both key academic participants (Merryman
Trubek and Galanter 1974) and a former Ford Foundation official (Gardner
) declared the program a failure, and support quickly evaporated.

he guiding assumption of the law and development movement was that law is
al to the development process. A related belief was that law was an instrument

that could be used to reform society and that lawyers and judges could serve as soci[illegible] engineers. As Merryman (1977) notes, not everyone subscribed to this view. A fe[illegible] participants in the movement argued that only minor changes could be effect[illegible] through legal reforms, and others contended that law reform should follow broad[illegible] changes in society, that is, that the proper aim of reform was to adjust the leg[illegible] system to social and economic changes that had already taken place. But the dom[illegible] nant view of law and development practitioners and theorists alike, although [illegible] unproven (Vorkink 1997), was that law reform could lead social change - that i[illegible] itself was an engine of change.

A second important belief was that educating the bench and bar in develop[illegible] countries would advance reform efforts. The gap between the law on the books [illegible] the law in action in developing countries was widely appreciated, and one of [illegible] solutions advanced was professional education (Burg 1977). It was thought th[illegible] lawyers and judges were properly educated about law's role in development [illegible] could be enlisted to close the gap. The idea was to turn members of both profess[illegible] into legal activists through education. Yet as one sympathetic chronicler of [illegible] movement observed, this idea was supported by nothing more than "hopeful sp[illegible] lation" that education could overcome values instilled by family, class, religion[illegible] other social forces (Lowenstein 1970:246-47).

The postmortems on law and development identify a number of pitfalls th[illegible] vocates of judicial reform ought to bear in mind (Burg 1977; Gardner 1980; Mer[illegible] man 1977; Trubek and Galanter, 1974). One is that the movement lacked any th[illegible] of the impact of law on development. Practitioners thus had no way to pr[illegible] reforms or predict the effects of various measures. A second failing was to[illegible] participation by the lawyers and others in the target country who either woul[illegible] to carry out the reforms or would be affected by them. Foreign legal consu[illegible] through a combination of expertise and access to funding, were often able to d[illegible] the content and pace of reform. A third problem was that the movement focused [illegible] the formal legal system to the exclusion of customary law and the other [illegible] ways in which many people in developing nations order their lives (Trubek [illegible] Galanter 1974).

But perhaps the most significant reason for its failure was the naive belief th[illegible] American legal system (and the legal culture generally), which Trubek and Ga[illegible] (1974:1062) refer to as "liberal legalism," could be easily transplanted to develop[illegible] countries. In the United States judges play a significant role in policymaking. [illegible] a result, lawyers are often able to engineer significant changes in policy thr[illegible] litigation. This is not true in civil law systems or indeed even in the United King[illegible] and other nations that share the same common law background as the United Sta[illegible] As Merryman (1977:479) put it, the law and development movement reflected [illegible] American "legal style," and this was a style that those in other cultures did no[illegible] particularly attractive.

At a 1995 conference hosted by the British Council, participants debated whether ie mistakes of the law and development movement are likely to be repeated. Faundez 997:13) argued that although the old programs and the Bank's new initiatives ap-ear to be quite similar on the surface, the context in which the Bank's current pro-ams are being carried out is significantly different. Behind the law and development iovement was the premise that the state "would initiate and promote the process of onomic development." By contrast, today the Bank sees law as facilitating market ansactions by defining property rights, guaranteeing the enforcement of contracts, id maintaining law and order. Because the state is no longer the protagonist of social iange, as in the law and development model, there is less room for error.

Yet as his analysis proceeded, Faundez seemed to be less sure that the mistakes of ie law and development movement would be avoided. He recognized that current evelopment theories, inspired by the work of Douglass North and other neo-insti-ional economists, still contemplate a role for the state. It is, to be sure, a different ne from the activist theory implicit in the law and development movement, and it one that is informed by economic analysis. "But it is unlikely that by shifting the ocus of attention from legal institutions to economic analysis this new approach will ianage to avoid the problems which so frustrated and disappointed members of the iw and development movement" (Faundez 1997:14). His concern is that all the nanswered questions that lurked behind the law and development movement—the ole of law and the formal legal system in development, the relationship between law nd politics, and among democracy, authoritarianism, and development—still remain.

If Faundez is ultimately uncertain that the Bank will not repeat the mistakes of the iw and development movement, both McAuslan (1997) and Thome (1997) have o doubts that it will. McAuslan advances a series of reasons why this is likely to iappen. Like Faundez, McAuslan underlines the absence of any empirical data con-ecting reform with development and the consequent disagreement even among eformers over priorities and strategy. In a commentary on McAuslan's article, Thome goes a step further. He believes the problem is not so much a lack of empirical data as he failure to reflect the data that are available. He asserts that all law reform, and idicial reform in particular, rests on three premises: first, that development requires modern legal framework resembling that in the United States; second, that this nodel establishes clear and predictable rules; and third, that the model can be easily ansferred. Yet, he says, empirical research has refuted all three assumptions.

McAuslan is also critical of what he argues is the Bank's focus on law reform to acilitate market transactions. The emphasis should be on promoting good gover-nance and alleviating poverty. An efficient and equitable market economy requires well functioning state-run institutions that can curb the abuses likely to arise as the market economy develops. He fears that in emphasizing the role of the judiciary in fostering economic growth, these considerations will be pushed aside. He cites as an example land-grabbing by elites as property rights are defined and allocated by the

state. Without judicial and administrative bodies capable of curbing such behavior, income inequalities will be exacerbated, and political instability may result.

McAuslan and Thome are not entirely negative. They do note that some Bank projects reflect the lessons of the law and development movement. The Financial and Legal Management Upgrading Project (FILUP) in Tanzania, for instance, has involved local lawyers from the beginning, both in studying the legal system and in developing proposals for change. Even the legislative drafting project for China, which they say is premised on an extreme view of the importance of law in the development process, is training local lawyers in the skills necessary for market reform. But even though the failings of the earlier law and development programs may be clear, both critics assert that pressures to produce results quickly will work against the gradual and incremental approach to law reform warranted by our current state of knowledge about the relationship between law and development.

Integrating Judicial Reform with Informal Enforcement Mechanisms

Judicial reform aims to buttress the rule of law and assure entrepreneurs that contracts will be enforced. Yet other institutions within society perform these same functions. Accountants and auditors issue standards and render judgments evaluating the performance of various economic actors. Credit bureaus provide an incentive to consumers and businesses alike to observe contracts by disseminating information about those who have failed to perform. And the media and nongovernmental organizations are often the first to detect and publicize arbitrary or illegal actions by government officials. Judicial reform projects that build upon or enhance the operation of such informal enforcement mechanisms will yield greater returns than those that do not. At the least, designers of reform projects must take the presence of these informal mechanisms into account. Otherwise, as Greif (1997) warns, the projects could backfire.

A forthcoming study by Kranton and Swamy of the effect that the introduction of formal courts had on rural credit markets in India illustrates just how a judicial reform project can go awry. Before there were formal courts, moneylenders relied on informal enforcement mechanisms to ensure that clients paid their debts. Resort to these mechanisms was costly and usually required the acquiescence of local leaders. Entry into the moneylending business was thus retarded and competition lessened. Although lack of competition meant that interest rates remained high, it also gave lenders a cushion that allowed them to extend payment terms and otherwise accommodate debtors experiencing difficulties in meeting their obligations.

The introduction of courts by the colonial authorities brought new entrants into the market. Interest rates declined, robbing lenders of the financial cushion that had

lowed them to carry borrowers in bad times. When drought hit, lenders went to ourt to foreclose on farmers' land. Riots and widespread social unrest followed.

As Kranton and Swamy note, the lesson is not that judicial reform is never appropriate, but that where other markets, such as those for insurance and futures, are missing, care must be taken to ensure that judicial reform does not have unintended consequences. Fafchamps's (1996) study of contracting in Africa supports this view. He found that rigid compliance with the terms of a written contract was difficult, if not impossible, in developing countries. Their economies are simply subject to too many exogenous shocks for contracts to be strictly enforced, which is why informal contract enforcement mechanisms build in such flexibility.

The problem in every case comes in determining how judicial reform will affect informal enforcement mechanisms, for a general theory of informal mechanisms and their interplay with formal mechanisms has yet to be advanced (Ellickson 1991). What is known is that when formal systems are deemed illegitimate, as they were in Moslem Central Asia after the Soviet takeover in the 1920s, disputes will be directed away from the formal system (Massell 1968). Ellickson (1991) submits that the division of labor between formal and informal mechanisms is affected by the technical complexity of the issues involved. He found, for example, that in northern California disputes between neighboring ranchers about the cost of a fence, which raised simple questions of fact and technology, tended to be resolved informally. By contrast, disputes involving the allocation of groundwater supply, where the facts were difficult to ascertain and resolution of the contested issues involved complex technical questions of return flow and allocation during shortages, were more likely to be presented to a court for adjudication.

The informal enforcement mechanisms that have drawn the most attention are reputation-based systems that permit merchants to carry on extensive trading relations over time and space in the absence of a court system that could ensure contract performance. Greif (1989) describes a system used by traders in North Africa and the Mediterranean in the 11th century, and in a later paper (1997), he shows similar reputation mechanisms at work in settings as diverse as the Wisconsin lumber industry, the New York diamond trading business, long-distance commerce in medieval Europe, and parts of contemporary Asia, Africa, and Latin America. The common denominator in all these examples is that the gains from repeat dealings provide the incentive necessary to ensure performance. That is, the discounted present value of the earnings stream that can be realized from future transactions exceeds the one-time wealth increase realizable from breaching the current agreement (Klein 1985).

The incentive to maintain a good reputation operates in other settings besides merchant-to-merchant relations. Credit bureaus—business associations that exchange information about the payment history of their customers—count on consumers' desire to buy on credit in the future to assure payment of current obligations (Klein 1992). A similar principle is behind consumer testing laboratories, better business

bureaus, and other groups that market seals of approval or provide quality guarantees (Klein 1997). When these groups certify that a product or business meets a certain standard, they are providing a visible sign of good reputation that can be used to generate future sales.

Development itself can affect the mix of formal and informal mechanisms in an economy. According to Besley (1995), one of the reasons informal financial institutions such as rotating savings and credit associations continue to operate in the developing world is that they spend far less than do banks and other formal financial institutions to ensure that borrowers repay their debts. Because their borrowers typically come from the same village, these institutions can rely on group pressure and other informal methods to see that the loans are repaid. Besley predicts that these institutions will lose their comparative advantage as the number of close-knit communities declines with the changes brought by economic development and that they will ultimately be supplanted by formal firms.

Milgrom, North, and Weingast (1990) also stress how increases in the costs of operating an informal enforcement mechanism lead users to turn to the formal legal system. They model a reputation-based system that resembles the one devised by long-distance traders in medieval Europe. Enforcement depends on each trader determining whether the other party to the contemplated exchange has failed to honor a contract, or cheated, in the past. If the other party is a cheater, the honest trader refuses to exchange with him. The threat of a boycott deters cheating.

But traders incur costs in ascertaining the past history of those with whom they contemplate exchanging—costs that increase as the economy grows. The number of potential trading partners on which information must be gathered expands, and the number of queries rises as the number of potential exchanges increases. According to Milgrom, North, and Weingast argue that eventually the costs to traders will exceed the costs of operating a formal judicial system. Rising transaction costs, they note, explain why national courts replaced the law merchant system of informal enforcement in Europe during the late Middle Ages.

Although the value of their analysis is its explicit focus on transaction costs and how changes in these costs dictate the choice between an informal and formal enforcement mechanism, to the extent that their work implies that development always makes informal mechanisms more costly than formal mechanisms, it is misleading. Development lowers at least some of the costs involved in operating an informal enforcement mechanism. The use of faxes, computers, and other technologies, for example, reduces the costs of compiling and disseminating information about the credit history of consumers and businesses (Ellickson 1991). How these increases and decreases net out, however, remains to be explored.

The line between formal and informal mechanisms may be fuzzy. In some cases a hybrid system appears. For example, the courts may enforce social customs or practices sanctioned by merchants (Benson 1989). Or bankers may hold titles to farmers'

l while their loans are outstanding, as in Thailand (Siamwalla and others 1993) Honduras (Stanfield and others 1990). Although actual possession gives the ks no formal legal right to foreclose on the land in case of default, bankers con-r the leverage from holding the title to be a sufficient guarantee of repayment.

luch remains to be learned about the working of informal enforcement mecha-ns and their relationship to the formal legal system. But at least some of the ways hich judicial reform can build on or complement informal systems are already arent. Some are obvious. Projects should capitalize on the power of the media to ce reform efforts by providing as much information about the judicial system as sible. In Argentina, for example, FIEL (1996) has proposed releasing information ilarly about judicial caseloads, case backlogs, and other indicators of judicial ductivity.

t a minimum, reform measures should try to bolster or complement informal orcement mechanisms. In the case of reputation mechanisms, this could mean losing the identity of the parties to lawsuits, the status of cases in litigation, and disposition of closed cases, including the amount of any damages awarded. Go-further current laws need to be reviewed to be sure they pose no obstacles to the and inexpensive dissemination of truthful public information about firms and viduals.

Whenever reputation information circulates, there is the possibility for abuse. se and defamatory material may be disseminated, jeopardizing privacy interests compromising opportunities to make a fresh start. There are, however, many s to strike a balance between these interests and the interests served by a well-ctioning reputation system. In the United States, the Federal Fair Credit Report-Act provides one model for balancing debtor and creditor interests. Another is ols juridically sanctioned process for publicizing information about those who e failed to pay their debts (World Bank 1997).

Often enforcing reputation mechanisms depends on an agreement among the ticipants to boycott anyone who has a history of breaching an agreement (Milgrom, rth and Weingast 1990). But some U.S. courts have ruled that a boycott is illegal en one or more of the firms participating in the boycott is a competitor of the is boycotted (ABA 1997). These rulings have been sharply criticized for misappre-ding the nature of anticompetitive agreements (Heidt 1986), but their influence still be reflected in competition law. Accordingly, a program to foster informal ract enforcement mechanisms should also encompass a review of the applicable npetition law.

On the basis of research in Ghana, Fafchamps (1996) recommends helping local siness communities develop systems to share information about the reliability of pliers and customers. Klein's (1992) analysis of credit bureaus shows that free-er problems and other market failures are endemic in the start-up phase and can vent the formation of credit reporting entities. In the United States credit bu-

reaus began as small, nonprofit associations, often run as an adjunct to the local chamber of commerce. Members were drawn from a tight-knit group of local merchants, and social ties supplied the incentives to overcome the problems of market failure. Where such incentives are missing, alternative ways of fostering the growth of credit reporting agencies should be considered. In Taiwan (China), the government's check clearinghouse serves as a substitute for a private credit bureau, charging a small fee for information about individuals who have bounced checks (Winn 1994).

Several other measures can be included in judicial reform projects to complement informal enforcement. One possibility is to transfer the responsibility for matters such as the registration of property rights to administrative agencies. This transfer can foster hybrid enforcement mechanisms. For example, technology now permits the creation of essentially paperless registries for land and other types of property. But paper titles may serve a purpose by permitting the development of a type informal mortgage based on physical possession of the certificate.

Conclusion

Although judicial reform projects are an accepted part of the development [illegible] scape, crafting an effective project poses several challenges. Many questions ab[illegible] how judicial reform affects the economy, and society generally, remain to be [illegible] swered. Not surprisingly, one result is that no consensus on the prerequisites [illegible] successful project has emerged. Accordingly, in the absence of a better theoret[illegible] understanding of the impact of judicial reform, care is required in designing [illegible] implementing projects (Greif 1997, Dakolias 1996; World Bank 1995, [illegible]. Reform must be preceded by an in-depth analysis of country needs [illegible] an analysis [illegible] must be continually reviewed as implementation proceeds

Notes

Richard E. Messick, a U.S.-trained lawyer, is a senior public sector management specialist [illegible] Public Sector Group of the Poverty Reduction and Economic Management Network of the [illegible] Bank.

1. An exhaustive list of potential interventions appears in Dakolias (1996). Other usef[illegible] sions of the range of possible reforms include the articles collected in Rowat, Malik, and [illegible] (1995).

2. Although *Rechtsstaat* is frequently rendered into English as "rule of law," as Fletcher [illegible] notes, this translation can be highly misleading. The word "law" in the English phrase [illegible] can mean either positive law, that is, any law enacted by a duly constituted government [illegible] law, that is, precepts that meet some test of morality and justice. The German *Recht* is [illegible] meaning to the latter than to the former, and thus a more accurate translation of *Rechtsstaat* [illegible] be "state ordered or ruled by natural law or justice." Similar translation problems arise with [illegible] European languages that, like German, have different terms for positive and natural law

ton, D.C. Processed. Spanish version: "Marco Institucional y Desarrollo Económico: La Reform. Judicial en América Latina." In Emundo Jarquin and Fernando Carrillo, eds., *La Economía Polít de la Reforma Judicial.* Washington, D.C.: Inter-American Development Bank.

Fafchamps, Marcel. 1996. "The Enforcement of Commercial Contracts in Ghana." *World Develo ment* 24:427–48.

Faundez, Julio. 1997. *Good Governance and Law: Legal and Institutional Reform in Developing Co tries.* New York: St. Martin's Press.

FIEL (Fundación de Investigaciones Económicas Latinoamericanas). 1996. *La Reforma del Poder Ju dicial en la Argentina.* Buenos Aires.

Fletcher. George P. 1996. *Basic Concepts of Legal Thought.* New York: Oxford University Press.

Fortescue, John. [1471?] 1979. *De Laudibus Legum Angliae [Praises of the Laws of England].* Trans lated by S. B. Chrimes. Reprint. Westport, Conn.: Hyperion Press.

GAO (U.S. General Accounting Office). 1993. *Foreign Assistance: Promoting Judicial Reform Strengthen Democracies.* Report GAO/NSAID-93-149. Washington, D.C.

Gardner, James. 1980. *Legal Imperialism: American Lawyers and Foreign Aid in Latin America.* Ma son: University of Wisconsin Press.

Greif, Avner. 1989. "Reputation and Coalition in Medieval Trade. Evidence on the Maghribi T ers." *Journal of Economic History* 49:857–82. Reprinted in Klein 1997.

———. 1997. "Contracting, Enforcement, and Efficiency: Economics beyond the Law." In Mic Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics* Washington, D.C.: World Bank.

Hammergren, Linn A. 1998. "Institutional Strengthening and Justice Reform." USAID Center Democracy and Governance, Washington, D.C. Processed.

Hayek, Friedrich A. 1960. *The Constitution of Liberty.* Chicago: University of Chicago Press.

Heidt, Robert. 1986. "Industry Self Regulation and the Useless Concept of 'Group Boyc *Vanderbilt Law Review* 39:1507–98.

Hirschman, Albert O. 1994. "The On-and-Off Connection between Political and Econ Progress." *American Economic Review* 84:343–48.

Hobbes, Thomas. [1651] 1962. *Leviathan.* New York: Collier Books.

IDB (Inter-American Development Bank). 1994. "Modernization of the State." Washington Processed.

———. 1995. "Frame of Reference for Bank Action in Programs for Modernization of the S and Strengthening of Civil Society." Washington, D.C. Processed.

Klein, Benjamin. 1985. "Self-Enforcing Contracts." *Journal of Institutional and Theoretical Ec ics* 141:594–600.

Klein, Daniel B. 1992. "Promise Keeping in the Great Society: A Model of Credit Inform Sharing." *Economics and Politics* 4:117–36. Reprinted in Klein 1997.

———, ed. 1997. *Reputation: Studies in the Voluntary Elicitation of Good Conduct.* Ann Ar University of Michigan Press.

Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-C Tests Using Alternative Institutional Measures." *Economics and Politics* 7:207–27.

Kranton, Rachel E., and Anand V. Swamy. Forthcoming. "The Hazards of Piecemeal Reform: B ish Civil Courts and the Credit Market in Colonial India." *Journal of Development Econom*

Lawyers Committee for Human Rights and Venezuelan Program for Human Rights and Educ and Action. 1996. *Halfway to Reform: The World Bank and the Venezuelan Justice System.* N York: Lawyers Committee for Human Rights.

nenstein, Steven. 1970. *Lawyers, Legal Education, and Development: An Examination of the Proces. of Reform in Chile*. New York: International Legal Center.

Auslan, Patrick. 1997. "Law, Governance, and the Development of the Market: Practical Problems and Possible Solutions." In Faundez 1997.

cLean, Roberto. 1996. "The Culture of Service in the Administration of Justice." *Transnationa, Law & Contemporary Problems* 6:139–64.

rtinez Neira, Nestor H. 1996. "The Ten Sins of Judicial Reform . . . and Some Anathemas." Paper prepared for Tinker Forum on the Role of the State in Latin America and the Caribbean Cancun, Mexico. Processed.

ssell, Gregory J. 1968. "Law as an Instrument of Revolutionary Change in a Traditional Milieu. The Case of Soviet Central Asia." *Law and Society Review* 2:179–214.

uro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110:681–721.

rryman, John H. 1977. "Comparative Law and Social Change: On the Origins, Style, Decline, and Revival of the Law and Development Movement." *The American Journal of Comparative Law* 25:357–83.

rget, Barry. 1997. "An Asian Perspective on Legal Modernization." Asian Development Bank. General Counsel's Office, Manila. Processed.

lgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchants, Private Judges, and the Champagne Fairs." *Economics and Politics* 2:1–23. Reprinted in Klein 1997.

rth, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge, K: Cambridge University Press.

z Perdomo, Rogelio. 1993. "Justice in Times of Globalization: Calls and Prospects for Change." *Justice and Development in Latin America and the Caribbean*. Washington, D.C.: Inter-American Development Bank.

or, Katharina. 1995. "Law Meets the Market: Matches and Mismatches in Transition Economies." Background paper prepared for *World Development Report 1996. From Plan to Market*. Processed.

ner, Richard. 1995. *Overcoming Law*. Cambridge, Mass.: Harvard University Press.

———. 1998. "Creating a Legal Framework for Economic Development." *The World Bank Research Observer* 13:1–13.

ka, Malcolm, Waleed H. Malik, and Maria Dakolias. 1995. *Judicial Reform in Latin America and the Caribbean*. Technical Paper 280. Washington, D.C.: World Bank.

wood, Robert. 1995. "Judicial Systems and National Economic Performance." Paper delivered at the Inter-American Development Bank's Second Annual Conference on Justice and Development in Latin America and the Caribbean, Montevideo, Uruguay. Processed. Spanish version: "Sistema Judicial y Desarrollo Económico." In Edmundo Jarquín and Fernando Carrillo, eds., *La Economía Política de la Reforma Judicial*. Inter-American Development Bank.

wood, Robert, Geoffrey Shepherd, and Celso Marcos de Souza. 1994. "Judicial Systems and Economic Performance." *The Quarterly Review of Economics and Finance* 34 (Special Issue):101–

ata, Ibrahim F. I. 1995. *The World Bank in a Changing World. Selected Essays and Lectures*. Vol. The Hague: Martinus Nijhoff Publishers.

wwalla, Ammar, and others. 1993. "The Thai Rural Credit System and Elements of a Theory. Public Subsidies, Private Information, and Segmented Markets." In Karla Hoff, Avishay Braverman, and Joseph E. Stiglitz, eds., *The Economics of Rural Organization: Theory, Practice, and Policy*. New York: Oxford University Press.

Smith, Adam. [1755] 1980. Unpublished manuscript quoted in Dugald Stewart, "Account of the Life and Writings of Adam Smith." In Adam Smith, *Essays on Philosophical Subjects.* Edited by W. P. D. Wightman and J. C. Bryce. Oxford: Clarendon Press.

Solow, Robert M. 1995. Review of *Trust: The Social Virtues and the Creation of Prosperity,* by Francis Fukuyama. *The New Republic* (September 11): 35–39.

Stanfield, David, Edgar Nesman, Mitchell Seligson, and Alexander Coles. 1990. "The Honduras Land Titling and Registration Experience." University of Wisconsin, Land Tenure Center, Madison. Processed.

Thome, Joseph R. 1997. "Comment on McAuslan's 'Law, Governance and the Development of [illegible] Market: Practical Problems and Possible Solutions.'" In Faundez 1997.

Trubek, David M. 1972. "Max Weber on Law and the Rise of Capitalism." *Wisconsin Law Review* 1972:720–53.

Trubek, David M., and Marc Galanter. 1974. "Scholars in Self-Estrangement: Some Reflections [illegible] the Crisis in Law and Development." *Wisconsin Law Review* 1974:1062–1101.

USAID (U.S. Agency for International Development). 1993. "Comments on the GAO Draft Report [illegible] Foreign Assistance: Promoting Judicial Reform to Strengthen Democracies." In GAO 1993.

Vargas, Juan Enrique. 1996. "Lessons Learned: Introduction of Oral Process in Latin America [illegible] Madeleine Crohn and William E. Davis, eds., *Lessons Learned.* Williamsburg, Va.: National Center for State Courts.

Vorkink, Andrew N. 1997. "The World Bank and Legal Technical Assistance: Current Is[illegible] World Bank, Legal Department, Washington, D.C. Processed.

Walker, William G. 1995. "Justice and Development: A Study." Paper prepared for the Administrator of the U.S. Agency for International Development (USAID), Washington, D.C. Processed.

Weder, Beatrice. 1995. "Legal Systems and Economic Performance: The Empirical Evidence [illegible] Rowat, Malik, and Dakolias 1995.

Williamson, Oliver E. 1995. "The Institutions and Governance of Economic Development [illegible] Reform." In Michael Bruno and Boris Pleskovic, eds., *Proceedings of the Annual World [illegible] Conference on Development Economics 1994.* Washington, D.C.: World Bank.

Winn, Jane K. 1994. "Relational Practices and the Marginalization of Law: Informal Fin[illegible] Practices of Small Businesses in Taiwan." *Law and Society Review* 28:193–232.

World Bank. 1992. *Governance and Development.* Washington, D.C.

———. 1994. *Governance: The World Bank's Experience.* Washington, D.C.

———. 1995. "The World Bank and Legal Technical Assistance: Initial Lessons." Policy Re[illegible] Working Paper 1414. Legal Department, Washington, D.C. Processed.

———. 1997. *World Development Report 1997: The State in a Changing World.* New York: [illegible] University Press.

The Costs and Benefits of Regulation: Implications for Developing Countries

J. Luis Guasch • Robert W. Hahn

he literature on the benefits and costs of regulation demonstrates that this issue can be plored systematically using standard economic analysis It also shows that regulation n have a significant adverse impact on economic growth. Specifically, regulation aimed t ontrolling prices and entry into markets that would otherwise be workably competitive kely to reduce growth and adversely affect the average standard of living. In addition, ness regulation can impose a significant cost on the economy. Nonetheless, social regu- ions may have significant net benefits for the average consumer if designed judiciously.

here are several policies developing countries might consider adopting to improve ir general approach to regulation The appropriate regulatory tools and framework d depend on many factors, including bureaucratic expertise, resource availability, po- al constraints, and economic impacts There is a general need to enhance the capabil- in evaluating regulation at the local and national levels

e past two decades have witnessed two trends in regulation. First, there has been unparalleled rise in new regulations related to health, safety, and the environ- nt. Second, there has also been substantial economic deregulation of certain in- tries in some countries, including airlines, trucking, railroads, financial markets, gy and telecommunications. At the same time, to complement the objectives of far reaching privatization programs, developing countries have begun to exam- egulations that keep prices inefficiently high.

he increased interest in regulatory reform can be explained in part by a growing erstanding of the impacts of regulation. The rationale that certain sectors had to egulated because they were "natural monopolies" vital to national social or stra- c interests is no longer considered valid. Moreover, the costs that regulations can ose on the economy are now better understood. Indeed, scholars now appreciate at regulation is subject to political influences and is rarely implemented with the sole purpose of improving economic efficiency; in many cases regulation has had

World Bank Research Observer

adverse effects on the economy. That argument forms the basis for the trend toward regulatory reform as globalization increases the pressure to reduce production costs and as officials react to the increased mobility of capital and labor by adjusting their policies to reflect the likely impact of regulations on price changes.

Not all regulation is on the decline, however. Citizens in many countries express a desire for more regulation in areas such as environmental protection, public health and safety standards. Rising incomes partly explain the increased interest; as consumers become wealthier, they demand more amenities such as cleaner air and water and better sanitation. And as politicians seek to supply more of these goods and services, they will also explore more efficient ways of supplying them.

Current political concerns with limiting tax increases in many countries are also creating incentives to use certain kinds of regulation. When legislators rein in spending and tax levels, regulation can be a useful substitute for achieving political objectives, such as redistributing income to particular interest groups in exchange for political support. In this kind of political environment, legislators adopt regulatory requirements or mandates whose costs are not directly paid for by taxpayers. Although less visible, these costs are nonetheless real. From the government's perspective, the effort appears to be relatively low cost. The federal budget is barely affected when a major change is mandated by regulation.

Why Regulate?

The most common economic arguments for regulatory intervention are market failure and considerations of equity. In the case of social regulation, a primary rationale is that without government intervention, individual companies may not take into account the full social cost of their actions. A firm may pollute excessively unless it incurs some implicit or explicit cost for polluting, for example, or workers may not have adequate information on health and safety hazards in the workplace to make fully informed choices. The argument for economic regulation has to do with the potential for improving production efficiency. If economies of scale exist, a single firm may, in theory, be able to produce more efficiently than several competing firms, but its monopolistic power may need be restrained through regulation. There is some justification for pursuing these objectives, but experience suggests that such rationales are often not persuasive in practice.

Correcting market failures and ensuring equity are laudable goals, but achieving those goals through regulation is not always successful. Just as there is potential for many kinds of market failure, so too is there potential for government failure. Economic regulation involves an understanding of the cost and demand structure of the industry, but a regulator typically does not have access to such information. Such

larly, health, environmental, and other social regulations must frequently be based on very limited information.

Political problems also lead to inefficient economic results. Because regulation redistributes resources and rents, politicians often use it to secure political gains rather than to correct market failures. A large array of regulatory instruments, such as quotas, licenses, and subsidies, is used to channel significant amounts of wealth to influential groups in society. In the United States, for example, price supports on peanuts resulted in an average annual consumer-to-producer transfer of $225 million (in 1987 dollars) with an associated deadweight loss of $34 million (Rucker and Thurman 1990). Wealth transfers are also a consideration in social regulation. Environmental and energy mandates frequently carry a heavy price tag.

Of course, if regulation becomes very inefficient and visible, pressure for reform ~w build. Firms with new technologies may lobby for reduced regulation; consum- and businesses may find ways to buy products and services at lower prices and opt t of the regulated markets. These considerations are particularly apt when demand 1 be met by tradable goods. Then the pressure to deregulate will come from do-stic producers who must compete with less regulated imports. In addition, pro-cers of tradable goods that rely on heavily regulated suppliers will have an interest facilitating deregulation of these sectors to lower their overall production costs.

stimating the Impact of a Regulatory Change

rhaps the most difficult task in estimating the impact of a regulatory change is solving the counterfactual: What would have happened in the absence of that ange? By comparing the effects of the counterfactual with the change induced by e regulation, it is possible to estimate the differences in costs and benefits between e two conditions and to calculate the impact on producers and consumers.

Once a counterfactual has been specified, there are five general approaches to imating the cost of regulation — econometric analysis, expenditure evaluation stud- , engineering cost analysis, productivity studies, and general equilibrium analysis.

- *Econometric studies* typically evaluate output markets directly or use production and cost functions to measure the impact of regulatory change. Although such studies do provide a formal statistical apparatus with which to test hypotheses, their formulation is typically quite general, glossing over the precise nature of actual production functions. Macroeconomic models are sometimes used in conjunction with econometric estimation to assess the economywide effects.
- *Expenditure evaluations* frequently rely on surveys of firms or businesses to determine costs of compliance. Direct surveys produce easily quantified (and often large) estimates of the cost of regulation, but such surveys face several

problems. The first involves potential respondent biases. For example, a firm or corporation may inflate its estimated costs in hopes that, if others follow suit, politicians will consider providing regulatory relief. More important, however, direct expenditure studies do not specify a counterfactual. For example, an automobile company may choose to install stronger bumpers on its cars even without a regulation forcing it to do so. Attributing the added cost of such bumpers to government regulation overstates the impact of regulation.

- *Engineering approaches* calculate the added cost of installing equipment directly, adjusting for quality changes. Again, the question is what kind of car would have been built in the absence of specific environmental regulations.
- *Productivity studies* chart the difference between observed productivity change over time and those that would have occurred in the absence of one or more federal regulations. These studies suffer from several problems, such as the reliance on expenditure data and an inability to specify the determinants of macroeconomic performance over time.
- *General equilibrium models*, which have become more popular recently, examine how a perfectly competitive market responds to a new policy, such as a change in regulation. The effects of a regulation can be linked to changes in output, employment, and in some cases welfare. Although general equilibrium models are not without their problems, including substantial data requirements, their results provide a better picture of regulatory effects in some cases. Simply stated, the methodological issue boils down to defining the conditions under which it is reasonable to assume away all but the most important effects.

The two basic approaches to measuring benefits rely either on asking people what they are willing to pay for changes in regulatory standards (contingent valuation) or on inferring from observed behavior the amount individuals actually pay for such quality changes. Although contingent valuation is particularly useful when markets do not exist for the commodity to be valued (for example, wilderness areas), it suffers from a likely divergence between what people choose to tell the interviewer and how they would behave under actual, rather than hypothetical, conditions.

Researchers rely on studies of averting behavior and on hedonic (shadow) price or wage methods to infer a willingness to pay. For example, certain neighborhoods have constructed barriers to mitigate the effects of highway or airport noise. The benefits of the noise reduction are then assumed to be at least the cost of the expenditure. Shadow price or wage methods attempt to evaluate the marginal value of quality improvements in specific amenities. For instance, the value workers implicitly place on safety is assumed to be the wage premium received by those working in more hazardous, although otherwise identical, circumstances. As such, econometric analyses of the implicit wage (or price) premiums can reveal the amount workers are willing to pay for improved workplace safety and, in the aggregate, their willingness to

y to prevent an expected fatality. The technique has been fruitfully applied in a mber of settings, including the valuation of reduced crime, reduced highway or port noise, cleaned-up hazardous waste sites, and other location-specific amenities. Hedonic estimation procedures are useful but rely on very indirect methods that ı, under certain circumstances, lead to identification problems. For instance, speci-ng all the relevant demand and supply characteristics that determine where people nose to live is a daunting econometric task, and one that may be severely biased if s of the determinants have been omitted. Moreover, in the case of estimating the ue of improved visibility and health, statistical problems often arise. Further, people ıy not be completely informed about certain risks, such as those associated with ıticular jobs, hazardous waste sites, and polluted air. Despite these problems, rapid vances in this relatively new technique promise improved empirical estimates of mmodities not explicitly traded in the marketplace.

imates of Benefits and Costs

the first study to synthesize data on the costs and benefits of regulation, Hahn and rd (1991) distinguished between transfer costs and efficiency costs. Transfers rep-sent payments from one group to another (for example, producers to consumers); iciency costs represent net losses in producer and consumer surpluses. Both mea-res are important, but for different reasons. Transfer payments provide a measure the winners and losers from regulatory change, while changes in net surplus pro-le an indication of the overall impact of a regulation on the economy or industry der investigation.

Tables 1 and 2, which show estimates of the costs of economic regulation and the sts and benefits of social regulation in the United States, demonstrate that it is ssible to explore systematically the costs and benefits of regulatory activity using ndard economic analysis. According to the analyses summarized in the tables, the iciency costs appear to be much smaller than the transfer costs—information that uld be taken into account when considering the effects of regulatory interven-n. For systematic economic studies of federal regulations in the United States, see denbaum and Delina (1978), Litan and Nordhaus (1983); Hahn and Hird 991); Hopkins (1992); Winston (1993); and Office of Management and Budget 997).

Hopkins (1992) argues that the costs of process regulation are also substantial. ble 3 shows that the total cost of federal regulation in 1991 was estimated at $542 llion, or about 9.5 percent of gross domestic product (GDP), including transfers. ne largest component of that cost was process regulation—the $189 billion in an-al expenditures related to government paperwork requirements, primarily for tax mpliance. These costs do not necessarily represent efficiency costs, however; one ust consider all aspects of a tax system in evaluating its impact on efficiency. None-

Table 1. *Estimated Annual Costs of Economic Regulation in the United States, 1988*
(billions of dollars)

Sector	*Efficiency costs*	*Transfers*	*Sources*
International trade	17.3	85.6–110.6	Hufbauer, Berliner, and Elliot (1986)
Telecommunications	< 14.1	< 42.3[a]	Wenders (1987)
Agricultural price supports	6.7	18.4	Gardner (1987)
Air transport	3.8	7.7	Morrison and Winston (1986, 1989)
Rail transport	2.3	6.8[a]	Winston (1985)
Postal rates	-	4–12	President's Commission on Privatization (1988)
Milk marketing orders/ price supports	0.4–0.9	0.9–3.5	Ippolito and Masson; Buxton and Hammond (both cited in MacAvoy 1977)
Natural gas[b]	0.2–0.4	5.0	Loury (1983)
Barge freight	0.2–0.3	0.6–0.9[a]	Litan and Nordhaus (1983)
Davis-Bacon Act	0.2[a]	0.5	Thieblot (1975) (updated)
Credit	0.05–0.5	0.15–1.6[a]	Litan and Nordhaus (1983)
Ocean freight	0.05–0.08	0.15–0.22[a]	Jantscher (1975)
Total	45.3–46.5	172.1–209.5	

— Not available.

a. Figure estimated using 3:1 ratio of transfers to efficiency costs.

b. Cost of natural gas regulation is expected to approach zero as all price controls are lifted.

Source: Hahn and Hird (1991).

Table 2. *Estimated Annual Costs and Benefits of Social Regulation in the United States, 1988*
(billions of dollars)

Sector	*Costs*	*Benefits*	*Source*
Environment	55.4–77.6	16.5–135.8	Hazilla and Kopp (1990); Freeman (1990); Portney (1990)
Highway safety	6.4–9.0	25.4–45.7	Crandall (1988)
Occupational safety and health (OSHA)[a]	8.5–9.0	Negligible	Crandall (1988); Denison (1979); Viscusi (1983)
Nuclear power	5.3–7.6		DOE policy study (cited in Litan and Nordhaus 1983)
Drugs	< 1.5–3.0		Peltzman (1973)
Equal employment opportunity	0.9	-	Weidenbaum and DeFina (1978); Litan and Nordhaus (1983)
Consumer product safety	> 0.034		Executive Office of the President (1990) (administrative cost only)
Total	78–107.1	41.9–181.5	

— Not available.

a. OSHA, Occupational Safety and Health Administration; DOE, Department of Energy.

Source: Hahn and Hird (1991).

Table 3. *Costs of Federal Regulation in the United States, Selected Years*
(billions of 1991 dollars)

Regulations	*1977*	*1988*	*1991*	*2000*
Environmental regulation	42	87	115	178
Other social regulation	29	30	36	61
Economic regulation-efficiency	120	73	73	73
Process regulation	122	153	189	221
Subtotal	*313*	*343*	*413*	*533*
Economic regulation transfers	228	130	130	130
Total costs	540	473	542	662

[illegible] Hopkins (1992)

heless, their sheer magnitude suggests that reducing paperwork would dramatically improve efficiency.

Outside the United States, few studies have estimated the costs of regulation. In Australia the total cost of regulation was estimated to be 9–19 percent of GDP in 1986 (OECD 1996). Mihlar (1996) estimates that the costs of regulation in Canada amounted to 12 percent of GDP. Based on an assumed ratio between private compliance costs and spending on regulatory programs, he extrapolated national regulatory costs from federal and provincial administrative budgets. Although crude, the calculation gives a rough estimate of the size of the regulatory burden.

Because these cost estimates are often cited without careful analysis, several points about them are worth noting. First, the figures are highly uncertain and often incomplete. Where there are uncertainties in the data, these should be conveyed as fully as possible to policymakers. Second, the figures developed using this approach to cost estimation are likely to understate the total impact of regulatory costs because they do not include the adverse effect that regulation typically has on innovation. Third, as shown in table 4, the cost of regulation as a fraction of GDP is significant for countries where such estimates are readily available, ranging from 7 to 19 percent. The Organisation for Economic Co-operation and Development, using a country-based macroeconomic model, has estimated that regulatory reform could increase GDP in the long run by as much as 3.5 percent in the United Kingdom and by as much as 6 percent in France, Germany, and Japan (OECD 1997a).

The Adverse Impacts of Regulatory Intervention

Many studies have attempted to estimate the adverse impacts of regulatory atervention.

- Christainsen and Haveman (1981) examined the effect of regulation on labor productivity and concluded that more than 10 percent of the contraction in

Table 4. *Costs of Regulation and Gains from Deregulation*
(percentage of GDP)

Economy	*Costs of regulation*	*Projected benefits of further economic deregulation*	*Source*
Australia	9–19	5.5	Commonwealth (1996) (cited in OECD 1996); Industry Commission (199[illegible]
Canada	11.8	—	Mihlar (1996)
European Union	—	4.5–7.0	Emerson and others (1988)
Germany	—	0.3	Lipschitz and others (1989)
Japan	—	2.3–18.7	OECD (1997a)
Netherlands	—	0.5–1.1	Sinderen and others (1994) (cited in 1997a); van Bergeijk and Haffner [illegible]
United States	7.2–9.5	0.3	Hopkins (1992); Winston (1993)

— Not available.

Note: These numbers are underestimates of the effects of deregulation because the studies do not [illegible] sectors in which deregulation can be beneficial. Further qualifications and elaborations of these [illegible] available from the authors.

the growth of labor productivity in the mid 1970s was attributable to [illegible] expansion of federal regulations. Guasch (forthcoming), who also look[illegible] labor costs, found that job growth was more robust in countries with [illegible] flexible labor markets (those at the top of table 5) than in those with [illegible] controls. Although many other factors affect employment, there are [illegible] reasons to believe that flexible labor market policies are likely to [illegible] employment.

- In examining the long-term growth effects of regulation on eight industries [illegible] 1973 to 1987, MacAvoy (1992) found economywide losses of 1.5–2 perce[illegible] U.S. gross national product.
- Studies examining environmental, health, and safety regulations have [illegible] qualitatively similar impacts. For example, Jorgenson and Wilcoxen [illegible] found that the costs of pollution control were associated with a reductio[illegible] more than 2.5 percent in the U.S. gross national product from 197[illegible] to [illegible] Robinson (1995) concluded that environmental and occupational healt[illegible] safety regulations cumulatively reduced multifactor productivity in the [illegible] facturing sector by more than 10 percent from 1974–75 to 1985–86.
- Research on the relationship between regulation and output growth in [illegible] countries by Koedijk and Kremers (1996) concluded from an index of regul[illegible] intensity that countries with the least regulation enjoyed the highest growth [illegible] output per person. The measures the authors construct are admittedly [illegible] but they may serve as a proxy for the degree to which markets are regulated [illegible] different countries.

Table 5. *Labor Regulations*

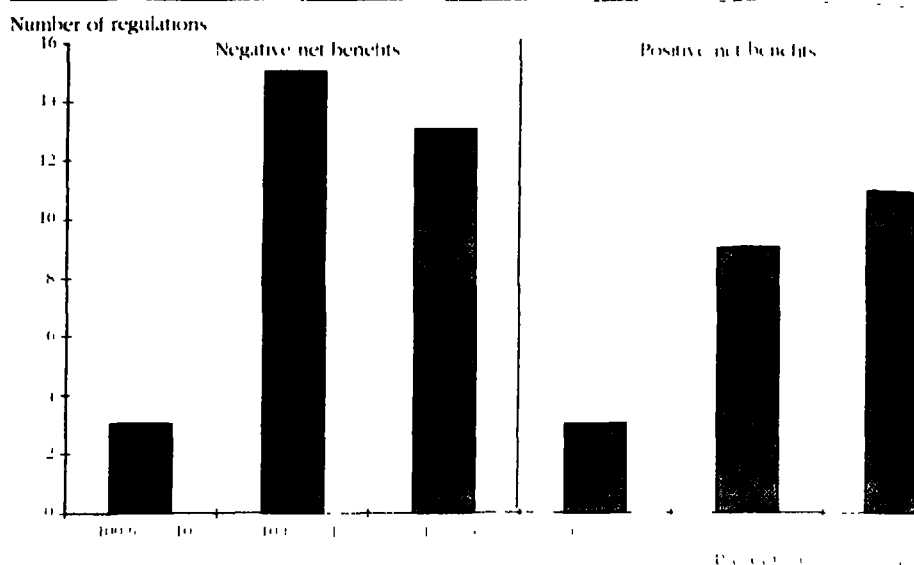
Country	*Payroll taxes (as a percent of the wage bill)*	*Severance payments*	*Collective bargaining*	*Employment growth, 1992–95*[b]	*Unemployment rate, 1996*
More flexible markets					
Australia	27.8	Low	Centralized	1.0	9.0
Chile	20.9	Low	Firm level	2.3	6.3
Japan	22.9	None	Firm level	0.6	2.5
Malaysia	24.3	Low	Firm level	3.3	2.8
New Zealand	11.5	None	Firm level	1.4	8.0
United States	20.1	None	Firm level	1.8	5.5
Less flexible markets					
Argentina	50.0[c]	High	Centralized	0.7	17.2
France	51.[illegible]	High	Centralized	0.4	11.6
[illegible]	52.8	High	Centralized	1.7	10.2
Spain	38.[illegible]	High	Centralized	1.6	22.4

a. The data for France, Spain, Italy, and Japan correspond to 1994, those for Malaysia to 1995, and those for Argentina and Chile to 1996. Severance payments are based on OECD indexes.
b. Employment growth is measured as annual average percentage growth.
c. Argentina amended its labor laws in 1996, and payroll taxes now average 43.0.
Source: Guasch (forthcoming).

Assessing the Gains from Regulatory Reform

In the area of social regulation, Hahn (1996), who reviewed more than five regulatory impact analyses (RIAs) covering health, safety, and environmental regulations from 1990 to mid-1995, reported considerable variation in the type and quality of analysis performed by government agencies. Cost-benefit analyses were often incomplete, and the costs of imposing the regulations were reported to be greater than the monetary benefits in more than 80 percent of the regulations. Based on this analysis, the net benefits from these regulations amounted to about $280 billion (1994 dollars) since 1990. Figure 1 shows the distribution of net benefits for 54 rules. The left side of the figure shows the number of rules with net costs that fall in various categories. The right side of the figure shows the number of rules with net benefits that fall in various categories. The aggregate net benefits are positive because a few of the rules have substantial benefits. Eliminating those that would not pass a benefit-cost test could increase the present value of net benefits by more than $[illegible]5 billion.

For various reasons, however, the RIA numbers cannot be taken at face value. Both theory and empirical evidence suggest that government agencies are likely to overstate substantially the aggregate net benefits of their programs (for instance, protecting the environment or improving safety in the workplace) to show that they are meeting the demands of interest groups.

Figure 1. *Distribution of Net Benefits of 54 U.S. Regulations, 1990 to mid 1995*
(billions of 1994 dollars)

Source: Hahn (1996).

A useful measure of the impact of regulations is how many lives are saved. Morrall (1986), in a review of several final and proposed regulations, calculates that the money needed to avoid each premature death varies over eight orders of magnitude—from roughly $100,000 to more than $5 trillion (1990 dollars).[1] This suggests that regulations could be developed that would prevent many more premature deaths while saving consumers money. Tengs and Graham (1996) found that reallocating regulatory expenditures within the United States to those investments that are most effective could avert an additional 60,000 deaths, or twice the current number. In addition, reallocating $8 billion (1994 dollars) in regulatory expenditures from the United States to developing countries could save more than 100 million additional lives (Hahn 1996).

The Benefits of Deregulation in Industrial Countries

The overall welfare gains from deregulation in the United States have been substantial. Aggregate welfare gains from eliminating entry and exit restrictions and freeing prices to market levels ranged from $35 billion to $46 billion (1990 dollars) a year, of which consumers gained $32 billion to $43 billion from lower prices and better services, and producers gained about $3 billion a year from increased efficiency.

[Ta]ble 6. Estimated Consumer and Producer Gains from Deregulation [in t]he United States, 1990

[Billi]ons of dollars)

[Industr]y	Consumers	Producers	Total	Further potential gains
[Air]lines	8.8–14.8	4.9	13.7–9.7	4.9
[Rail]ways	7.2–9.7	3.2	10.4–12.9	0.4
[Roa]d freight	15.4	(4.8)	10.6	0
[Tele]communications	0.7–1.6		0.7–1.6	11.8
[Cab]le television	0.4–1.3		0.4–1.3	0.4–0.8
[Stoc]k market trading	0.1	(0.1)	0	0
[Nat]ural gas				4.1
[Tot]al	32.6–43.0	3.2	35.8–46.2	21.6–22.0

Not available.

[Source]: Winston 1993.

[produ]ce[r] costs (table 6). Winston (1993) estimates that additional gains from eliminat[in]g remaining distortions could be more than $20 billion a year. Even so, there is [evi]dence that the gains are likely to be significantly understated. In a recent paper, [Wi]nston (1998) notes that although industry may adjust prices to reflect marginal [co]sts quickly after deregulation, it takes time to optimize production. He argues that [pol]icymakers and the public tend to notice only the short term effects and therefore [un]dervalue the benefits of deregulation. the positive impact that deregulation has on [inn]ovation is frequently overlooked. Such innovations increased productivity and [redu]ced operating costs by one fourth to more than one-half in different industries.

Several studies yield similar results on the adverse consequences of economic [reg]ulation. Caves, Christensen and Swanson (1981) compared the productivity [gro]wth of U.S. railroads from 1956 to 1974 with the growth of Canadian railroads [duri]ng the same period. Both industries had access to the same technology, but [Can]ada's railroads were not as heavily regulated as those in the United States. The [auth]ors argue that regulation substantially reduced productivity growth, estimating [that] had the U.S. railroads (with a growth rate of 0.5 percent) had experienced the [same] growth as Canada's (3.3 percent) the cost of providing rail services in 1974 [wou]ld have been $13.8 billion (1985 dollars) lower. Willig and Baumol (1987) esti[ma]ted that after rail deregulation in the United States, annual operating expenses [dro]pped 26 percent from 1980 to 1985, while traffic volume remained virtually [unch]anged and investment increased.

The empirical evidence on the trucking, airline, telecommunications, and finan[cial] industries is impressive.

- Average unit costs in the U.S. trucking industry declined from 30 cents a ton-mile in 1977 before deregulation to 10 cents a ton-mile in 1983 after

deregulation (in 1977 dollars). The annual welfare loss from regulation of ra[illegible]
and motor carrier rates was estimated at $1 billion to $4 billion (1977 dollar[illegible]
(Braeutigam and Noll 1984; Winston and others 1990).

- The airline industry reduced total costs per unit of service by approximately [illegible] percent; labor costs were cut as well (by 17 percent at American Airlines and [illegible] percent at United Airlines) with little effect on output in the first few [illegible] following deregulation (Caves and others 1987). In addition, excess capacity declined and productivity rose. Morrison and Winston (1995) estimate the [illegible] annual gains to passengers at $18.4 billion (1993 dollars).
- By 1996 long-distance telephone rates in the United States had fallen by more than 70 percent as a result of the divestiture of AT&T in 1984 (Taylor and [illegible] 1993; *Wall Street Journal* 1991). The emergence of profitable services such [illegible] cellular telephony and voice messaging after divestiture shows how regulation can slow the introduction of new products and discourage innovation. Although the concept of cellular phones was discussed in the late 1940s and the technology was available in 1973, the Federal Communications Commission did not be[illegible] to issue licenses until ten years later—a delay that, by one estimate, cost the U[illegible] economy more than $25 billion a year in 1983 (Rohlfs, Jackson, and Ke[illegible] 1991), or about 2 percent of gross domestic product. Similarly the delay [illegible] introducing voice messaging services cost more than $1.3 billion (1994 dollars) a year (Hausman and Tardiff 1996).
- Postderegulation effects have been observed in the securities, investment [illegible] banking sectors. For example, when brokerage fees were deregulated [illegible] dropped by 25 percent and savings from overall consolidation and cost red[illegible] tion amounted to 30 percent (Jarrell 1984). Studies have shown that even [illegible] accounting for changes in the services offered, the cost reductions were sig[illegible] cant. In the United States partial deregulation of the banking and savings [illegible] loan industries resulted in employment cuts of more than 20 percent dur[illegible] 1984–93 and an increase in productivity (as measured by revenue per employ[illegible] of more than 300 percent (Guasch and Spiller forthcoming).

Although the database outside the United States is less extensive, there is reas[illegible] believe that the gains from deregulation of many industries elsewhere could be [illegible] stantial (see table 4). For example, airline fares in Europe are roughly twice as exp[illegible] sive as in the United States (Airfare Management Unit 1995, 1996; Consulting S[illegible] vices Group 1995, 1996), but profitability is well below that of U.S. carriers. [illegible] price and entry restrictions could reduce fares and benefit consumers. Indeed, [illegible] high-cost carriers, such as Iberia and Air France (both state-owned), have survived only with government aid. Good, Röller, and Sickles (1993), who argue that liber[illegible] ization would lead to competition between international carriers and a converge[illegible] of cost structures, estimate that if the European airline industry were as efficient [illegible]

U.S. airline industry, it would have saved approximately $4 billion a year in
6 dollars.

eregulation of electricity markets in Europe also offers significant opportunities
gains (Electricity Association Services Ltd. 1996). In Germany, for example, strict
lations require domestic companies to purchase electricity from regional pro-
rs even though cheaper power is often available nearby. The extent of the poten-
gains for German consumers is difficult to estimate, but in the United Kingdom,
gy deregulation resulted in a 70 percent increase in productivity and an 18 to 21
ent reduction in franchise contract prices (OECD 1997a). Elsewhere in the
opean Union, firms pay over 50 percent more for their electricity than do their
erican counterparts. Moreover, the impact of higher energy prices on the overall
iomy can be quite significant (Navarro 1996). For example, a 30 percent in-
se in electricity prices tends to raise the prices of goods such as paper and pulp,
ls, chemicals, and glass by roughly 2.5 percent.

efits of Deregulation in Developing Countries

untries that have deregulated, the efficiency gains have been quite significant.
example, deregulation of entry into the long distance telephone market in Chile
cut rates by 50 percent, making them close to U.S. rates (Guasch and Spiller
hcoming). In some Latin American countries, private sector participation in the
communications sector has cut waiting time for installation of new lines from a
imum of two years to a matter of weeks. At the port terminals in Buenos Aires,
petition in operations has led to an 80 percent reduction in the fees charged.
l opening port operations to multiple parties in the port of Montevideo has in-
sed productivity by 300 percent. All those results were achieved within a year of
gulation (Guasch 1996).

A study of Argentina (Fundacion de Investigaciones Economicas Latinoamericanas
9) assesses the welfare cost of regulations and other government interventions in
980s (table). The total costs of regulation and state intervention amount to
e than $4 billion a year (1990 dollars), and this is only for the selective listed
ventions. Using a general equilibrium model, Chisari, Estache, and Romero
9) estimate the gains from privatization and regulation in Argentina at about
percent of gross domestic product, or $3.3 billion. They also find that *all* income
es benefit.

It would be useful to assemble data on regulatory costs in other developing coun-
s comparable to those assembled for Argentina. Yet there is no shortage of spe-
cases where economic regulation has had adverse consequences. For example,
uguayan firms and consumers are paying an implicit tax of at least 30 percent for
ter, phone, and electricity, thus hindering the competitiveness of Uruguay's prod-
s compared with those of Argentina, Brazil, and Paraguay, its fellow members in

Box 1. *Montevideo Taxicab Market*

Entry restrictions for taxicabs in Montevideo, Uruguay made the 1990 market price of a taxicab license $60,000 (in 1990 dollars). Although the license is nominally lower than the $125,000 price in New York, lower Uruguayan per capita income means that the market value of the license as a proportion of per capita income is more than four times higher. The regulation of the taxicab market has led to a scarcity of taxicabs, as reflected in difficulty in hailing taxicabs, in high costs borne by consumers, and in capture and wasteful rent-seeking activity by the taxi owners association.

Source: Guasch and Spiller (forthcoming).

As noted earlier, Mexico is reviewing regulations for major federal agencies. The purpose of the review is to eliminate unnecessary regulations, simplify regulations that are unnecessarily burdensome, and make the process more transparent (box 2). By the end of 1997, approximately half of all regulations (called formalities) had been reviewed in seven of twelve ministries. Of those reviewed, 38 percent were scheduled to be eliminated and an additional 54 percent were scheduled to be simplified in 1998.

Conclusions

Regulations receive relatively little scrutiny, both because politicians wish to hide the cost of regulation from citizens and because estimating the costs and benefits of regulation is difficult. A better information base on the economic impact that different regulations have would enhance public decisionmaking. There are several poli-

Box 2 *Regulatory Reform in Mexico*

The government of Mexico is undertaking an examination of regulatory structure at the federal, and local levels. The aims of the Agreement for the Deregulation of Business Activity include streamlining federal regulations, reducing corruption by codifying regulation, and helping to promote efficient and effective regulation. The program has enjoyed some early successes. Recent legislation simplifies administrative procedures, requires a quicker administrative response time, and reduces paperwork for foreign investors. In addition, a series of legal reforms aims to simplify court proceedings and reduce the costs of commercial lending. As a result of these reforms, Mexico City's Superior Court reports that the number of civil suits filed fell by 24 percent from 1995 to 1996. Reforms by agency rule simplification and elimination are also proceeding swiftly. For example, the approval time for a business to begin operation has been reduced from an average of more than 200 working days to a maximum of 21 working days. Finally, a complete inventory of federal rules in effect is available on the Internet; easy access should help to reduce corruption and compliance costs.

Source: Secretaría de Comercio y Fomento Industrial (1997).

cies that developing countries might consider; the recommendations here are pu
posely general. In that spirit, the first important point to make is that effective po
cies will differ across countries. The appropriate regulatory tools and framework w
depend on several factors, including bureaucratic expertise, resource availability p
litical constraints, and economic impacts.

There is a general need, moreover, to enhance the capability for evaluating regul
tion at the local and national levels (Hahn forthcoming), as illustrated by the absen
of even rudimentary data in many countries on the effects of regulation. Countri
should attempt to develop a "regulatory budget" that would show the econom
effects of regulations and that would be published along with the government's fis
budget. Such a capability will take time to develop.

Several jurisdictions, including some in developing countries, are putting proc
dures in place that would require a benefit-cost analysis before significant regulatio
could be implemented (OECD 1997b). This is likely to have a constructive influen
on public policy by providing better information and holding government offici
and political leaders more accountable (Hahn and Litan 1997). In the short term
is important for agencies charged with administering regulations to begin by assem
bling crude cost and benefit data. For example, an agency could specify the rational
for a proposed regulation, the likely direct and indirect costs, a qualitative descri
tion of benefits, an assessment of other alternatives, including the status quo, and
explanation of why other alternatives were not selected if they are likely to be bett
for the average citizen.

Such analyses should not be too burdensome. For regulations that have a limit
impact, no analysis may be necessary. For regulations with potentially large ec
nomic consequences, more resources should be devoted to evaluation. Ideally
analyses should be both prospective and retrospective, so that analysts can learn
to improve their impact assessments by comparing their predictions with actual
litical outcomes. To start with, we recommend developing a low cost informati
management system that highlights some of the more important economic impa
of regulation. Front-line agencies need to be involved in the process so that
become more sensitive to the economywide impacts of their proposals.

As administrative capabilities evolve, a more thorough cost-benefit analysis will
required to support regulatory reform. Because economic regulation often results
economic inefficiency, the burden of proof should be on those who wish to main
such regulation. In the case of social regulation, flexibility should be encourage
that consumers and producers are able to innovate in response to regulations.
for example, performance standards for meeting a pollution goal are generally pr
ferred to standards that dictate the use of a particular technology. Of course, the
amount of flexibility in a regulatory policy should be based, in part, on the ability of
the administrative agency to implement it effectively (Hartman and Wheeler 199

Although an economic analysis of regulatory policies can be helpful, regulations often have unexpected and perverse consequences (Ackerman and Hassler 1981). Thus it is better to proceed with extreme care and err on the side of less regulation, particularly when dealing with economic interventions. Where there is no clear economic rationale for a regulation, it should be removed (Hahn 1998). Licensing and price or quota interventions, for example, do not serve the public interest but instead transfer political favors to preferred constituencies (Huber and Thorne forthcoming; Guasch and Spiller forthcoming). Removing such distortionary favors may not be easy in many cases and may involve resource transfers to politically powerful constituencies.

As they consider reforms, policymakers need to give a great deal more thought to the design of regulatory frameworks. In some instances, partial deregulation may not lead to an improvement over the status quo. For example, removing price restrictions but retaining entry barriers could lead to inefficient pricing. Full deregulation can lead to problems with monopoly unless great care is taken in managing the transition to a deregulated environment. The point is that the strategy for regulatory reform is critical to its effectiveness. Another set of problems stems from a tendency for a single-mission agency (health or education, say) to consider its mandate exclusively and to overstate the benefits of its program and understate the costs. As noted above, one way to address this problem is to require the agency to develop more data on the costs of specific regulatory proposals. A second is to limit the agency's mandate. Other options include sunset requirements that would limit an agency's authority to a fixed period unless renewed by legislative mandate and requiring the approval of a central—independent—agency that is primarily concerned with the economywide impacts of regulations (Hahn 1997). Because officials are concerned about issues of equity and efficiency, the regulations they write tend to be unduly complicated. This complexity not only gives bureaucrats and lawyers control over decisions but also makes it difficult for average people to understand the economic implications involved. The more transparent the regulations, the more they are likely to reduce the potential for corruption and increase the perceived legitimacy of the system. Straightforward language makes careful scrutiny possible and limits the likelihood that political interest groups will capture the benefits. A move toward greater transparency will occur as people begin to understand some of the hidden costs of regulation.

Developing countries have begun to realize the benefits of reforming economic regulation, but much remains to be done in the area of social regulation. Yet, it is beginning to appear on the policy agenda, if not from domestic pressure, then from interest groups in industrial countries.

The overall lesson is not that regulation is generally undesirable but that it often has undesirable economic consequences. Moreover, these effects result partly from

political forces that lead to inequitable redistribution of wealth (Stigler 1971). We believe such forces can be mitigated by more sharply evaluating the consequences and tradeoffs involved before a regulatory policy is set in stone.

Notes

J. Luis Guasch is lead economist for the Latin America and the Caribbean Regional Office and the Development Economics Group at the World Bank and is professor of economics at the University of California, San Diego. Robert W. Hahn is director of the American Enterprise Institute (AEI)–Brookings Joint Center for Regulatory Studies, a resident scholar at AEI, and a research associate at the Kennedy School of Government, Harvard University. The authors thank Ed Campos, Michael Klein, Brian Levy, Joseph Stiglitz, and the referees for their helpful comments and Jonah Siskin, Lisa Taber, and Fumie Yokota for their excellent research assistance.

1. Hopkins' estimate for the total cost of regulation includes transfer costs and process costs. Subtracting transfer costs yields an estimate of $413 billion, or more than $1,500 per person in 1991. Process costs account for about half of the $1,500. For a critique of Hopkins' analysis, see Office of Management and Budget (1997).

2. Employment declined from 260,000 in 1987 to 190,000 in 1990.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Ackerman, Bruce A., and William T. Hassler. 1981. *Clean Coal/Dirty Air: Or How the Clean Air Act Became a Multibillion-Dollar Bail-Out for High-Sulfur Coal Producers and What Should Be Done About It.* New Haven, Conn.: Yale University Press.

Airfare Management Unit. 1995. *The American Express European Airfare Index.* Paris: American Express.

———. 1996. *The American Express European Airfare Index.* Paris: American Express.

Anderson, Robert C., Alan Carlin, Albert M. McGartland, and Jennifer B. Weinberger. 199[illegible]. "Savings from the Use of Market Incentives for Pollution Control." In Richard F. Kosobud and Jennifer M. Zimmerman, eds., *Market Based Approaches to Environmental Policy: Regulatory Innovations to the Fore.* New York: Van Nostrand Reinhold.

Braeutigam, Ronald R., and Roger G. Noll. 1984. "The Regulation of Surface Freight Transportation: The Welfare Effects Revisited." *Review of Economics and Statistics* 56 (February): [illegible]

Caves, Douglas W., Laurits R. Christensen, and Joseph A. Swanson. 1981. "The High Cost of Regulating U.S. Railroads." *Regulation* (January/February): 41–46.

Caves, Douglas W., Laurits R. Christensen, Michael W. Tretheway, and Robert J. Windle. [illegible] "An Assessment of the Efficiency Effects of U.S. Airline Deregulation via an International Comparison." In Elizabeth E. Bailey, ed., *Public Regulation: New Perspectives on Institutions and Policies.* Cambridge, Mass.: MIT Press.

Chisari, Omar, Antonio Estache, and Carlos Romero. 1997. "The Distribution of Gains from Utility Privatization and Regulation in Argentina." In "The Private Sector in Infrastructure: Strategy, Regulation, and Risk." World Bank, Finance, Private Sector and Infrastructure Network, Washington, D.C. Processed.

istainsen, Gregory B., and Robert H. Haveman. 1981. "Public Regulations and the Slowdown in Productivity Growth." *American Economic Review* 71(May):320–25.

nsulting Services Group. 1995. *The American Express Domestic Airfare Index.* New York: American Express.

—. 1996. *The American Express Domestic Airfare Index.* New York: American Express.

ndall, Robert W. 1988. "What Ever Happened to Deregulation?" In David Boaz, ed., *Assessing the Reagan Years.* Washington, D.C.: Cato Institute.

nison, Edward F. 1979. *Accounting for Slower Economic Growth: The United States in the 1970s.* Washington, D.C.: Brookings Institution.

ctricity Association Services Ltd. 1996. "International Electricity Prices," Issue 23. London.

erson, Michael, Michel Aujean, Michel Catinat, Philippe Goybet, and Alexis Jacquemin. 1988. *The Economics of 1992: The EC Commission's Assessment of the Economic Effects of Completing the Internal Market.* Oxford: Oxford University Press.

ecutive Office of the President. 1990. *Budget of the United States Government, Fiscal Year 1991.* Washington, D.C.: U.S. Government Printing Office.

eman, A. Myrick. 1990. "Water Pollution Policy." In Paul R. Portney, ed., *Public Policies for Environmental Protection.* Washington, D.C.: Resources for the Future.

ndacion de Investigaciones Economicas Latinoamericanas. 1991. *Regulatory Costs in Argentina.* Buenos Aires: FIEL/ANZA.

rdner, Bruce L. 1987. "Protection of U.S. Agriculture: Why, How, and Who Loses?" Department of Agricultural and Resource Economics Working Paper 87-15, University of Maryland, College Park, Md. Processed.

od, David H., Lars Hendrich Roller, and Robin C. Sickles. 1993. "U.S. Airline Deregulation: Implications for European Transport." *Economic Journal* 103(July):1028–41.

asch, J. Luis. 1996. "Lessons for Port Reforms." In J. Luis Guasch and Leandre Amargos, eds., *New Port Policies in Latin America and Caribbean.* Barcelona: New Press.

—. Forthcoming. *Labor Reform, Markets and Job Creation: The Unfinished Agenda in Latin American and Caribbean Countries.* Directions in Development Series. Washington, D.C.: World Bank.

asch, J. Luis, and Pablo Spiller. Forthcoming. *Managing the Regulatory Process: Design, Concepts, Issues, and the Latin America and Caribbean Story.* Washington, D.C.: World Bank Economic Development Institute.

hn, Robert W. 1996. "Regulatory Reform: What Do the Government's Numbers Tell Us?" In Robert W. Hahn, ed., *Risks, Costs, and Lives Saved: Getting Better Results from Regulation.* New York: Oxford University Press and AEI Press.

—. 1997. "Achieving Real Regulatory Reform." *University of Chicago Legal Forum* 1997:143–58.

—. 1998. "Government Analysis of the Benefits and Costs of Regulation." *Journal of Economic Perspectives* 12(4):201–10.

hn, Robert W., ed. Forthcoming. *Reviving Regulatory Reform: A Global Perspective.* Washington, D.C.: AEI Press and Brookings Institution.

hn, Robert W., and John Hird. 1991. "The Costs and Benefits of Regulation: Review and Synthesis." *Yale Journal on Regulation* 8(Winter):233–78.

hn, Robert W., and Robert E. Litan. 1997. *Improving Regulatory Accountability.* Washington, D.C.: AEI and Brookings Institution.

rtman, Raymond S., and David Wheeler. 1995. "Incentive Regulation: Market-Based Pollution Control for the Real World?" In Claudio R. Frischtak, ed., "Regulatory Policies and Reform: A

Comparative Perspective." World Bank, Private Sector Development Department, Washington, D.C. Processed.

Hausman, Jerry, and Timothy Tardiff. 1996. "Valuation and Regulation of New Services in Tele-communications." Discussion Paper. Massachusetts Institute of Technology, Economics Department, Cambridge, Mass. Processed.

Hazilla, Michael, and Raymond J. Kopp. 1990. "The Social Cost of Environmental Quality Regulations: A General Equilibrium Analysis." *Journal of Political Economy* 98(4):853–73.

Hopkins, Thomas D. 1992. "Costs of Regulation: Filling the Gap." Report prepared for the Regulatory Information Service Center, Washington, D.C. Processed.

Huber, Peter W., and John Thorne. Forthcoming. "Economic Licensing Reform." In Hahn (forthcoming).

Hufbauer, Gary C., Diane T. Berliner, and Kimberlie A. Elliot. 1986. *Trade Protection in the United States: 31 Case Studies.* Washington, D.C.: Institute for International Economics.

Industry Commission. 1995. "The Growth and Revenue Implications of Hilmer and Related Reforms." A Report by the Industry Commission to the Council of Australian Governments. Belconnen, Australian Capital Territory. Processed.

Jantscher, Gerald R. 1975. *Bread upon the Waters: Federal Aid to the Maritime Industries.* Washington, D.C.: Brookings Institution.

Jarrell, Gregg A. 1984. "Change at the Exchange: The Causes and Effects of Deregulation." *Journal of Law and Economics* 27(October):273–312.

Jorgenson, Dale W., and Peter J. Wilcoxen. 1990. "Environmental Regulation and U.S. [illegible] Growth." *RAND Journal of Economics* 21:314–40.

Koedijk, Kees, and Jeroen Kremers. 1996. "Market Opening, Regulation and Growth in Europe." *Economic Policy: A European Forum* 23(October):445–67.

Lipschitz, Leslie, Jeroen Kremers, Thomas Mayer, and Donogh McDonald. 1989. "The [illegible] Republic of Germany: Adjustment in a Surplus Country." Occasional Paper 64. International Monetary Fund, Washington, D.C. Processed.

Litan, Robert, and William Nordhaus. 1983. *Reforming Federal Regulation.* New Haven: [illegible] University Press.

Loury, Glenn C. 1983. "Efficiency and Equity Impacts of Deregulation." In Robert H. Haveman and Julius Margolis, eds., *Public Expenditure and Public Policy Analysis.* Boston: Houghton Mifflin.

MacAvoy, Paul. 1977. *Federal Milk Marketing Orders and Price Supports.* Washington, D.C.: American Enterprise Institute.

———. 1992. *Industry Regulation and the Performance of the American Economy.* New York: [illegible] Norton.

Mihlar, Fazil. 1996. "Regulatory Overkill: The Costs of Regulation in Canada." Fraser [illegible] Vancouver. Processed.

Morrall, John F. 1986. "A Review of the Record." *Regulation* 10(November–December) [illegible]

Morrison, Steven A., and Clifford Winston. 1986. *The Economic Effects of Airline Deregulation.* Washington, D.C.: Brookings Institution.

———. 1989. "Enhancing the Performance of the Deregulated Air Transportation System." *Brookings Papers on Economic Activity: Microeconomics* 61–123.

———. 1995. *The Evolution of the Airline Industry.* Washington, D.C.: Brookings Institution.

Navarro, Peter. 1996. "Electric Utilities: The Argument for Radical Deregulation." *Harvard Business Review* 74(1):12–25.

Winston, Clifford. 1985. "Conceptual Developments in the Economics of Transportation: An I terpretive Survey." *Journal of Economic Literature* 23(1):57–94.

———. 1993. "Economic Deregulation: Days of Reckoning for Microeconomists." *Journal of E nomic Literature* 31:1263–89.

———. 1998. "U.S. Industry Adjustment of Economic Deregulation." *Journal of Economic Pers tives* 12(3):89–110.

Winston, Clifford, Thomas M. Corsi, Curtis M. Grimm, and Carol Evans. 1990. *The Econo Effects of Surface Freight Deregulation.* Washington, D.C.: Brookings Institution.

World Bank. 1997. *World Development Report 1997. The State in a Changing World.* New Y Oxford University Press.

Road Funds, User Charges, and Taxes

Ken Gwilliam • Zmarak Shalizi

Insufficient or uncertain budgetary allocations to road maintenance have resulted in road deterioration that has significantly increased production and transport costs in many countries. To avoid this problem, highway professionals advocate the establishment of dedicated road funds, managed by independent road boards made up of user representatives. The road boards would have the power to determine both the level of charges for road use and the level of expenditure on road maintenance. By contrast, macroeconomists and public finance specialists have tended to resist the establishment of dedicated road funds. They argue that road funds reduce fiscal flexibility, do not adequately address problems associated with the provision of public goods or the internalization of externalities, and are often not well managed.

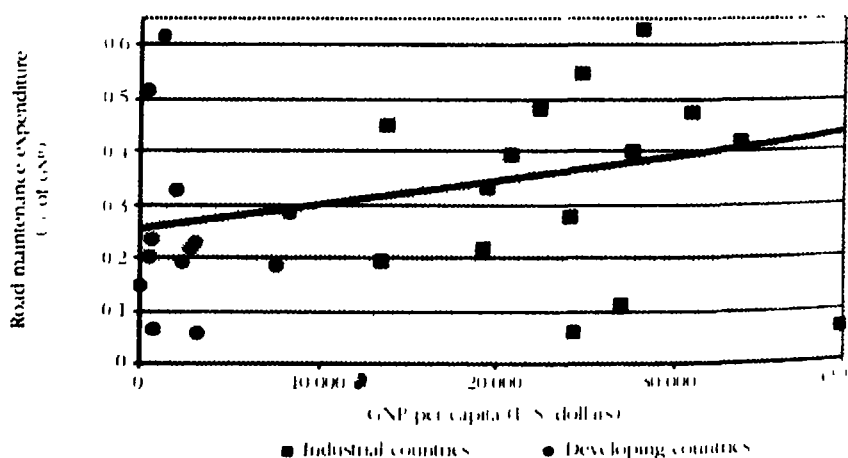
In general, there are two long-term institutional options for reconciling fiscal prudence with road maintenance: a road agency that is operated commercially (subject to the normal oversight of behavior accorded to privatized monopolies), or a reformed and well-functioning budget process. This article argues that road funds must be viewed as a provisional, case-specific intermediate step in the direction of one of the long-term solutions. The role and nature of road funds should be assessed not on general principles but on a case-by-case basis through the analysis of likely micro- and macroeconomic effects. The article recommends indicators for use in specific cases to determine whether a road fund should be introduced, continued, or abolished.

Spending on roads accounts for between 0.5 and 10 percent of public expenditure and between 10 and 20 percent of the development budget in many countries. Expenditures on road construction and rehabilitation, which are usually classified as capital expenditures, financed through borrowing and subsequent debt servicing, are not discussed in this paper. This article focuses on expenditures for periodic maintenance, routine maintenance, and operational management of roads. These costs are usually treated as current expenditures financed through annually budgeted allocations from the central treasury.

The World Bank Research Observer, vol. 14, no. 2 (August 1999), pp. 159–85

Road maintenance expenditures account for 30 to 60 percent of total road expenditures and as much as 0.5 percent of gross national product, or GNP (table A-1). Some weak evidence indicates that as road network densities and trade and traffic volumes increase with economic growth, road maintenance expenditures increase with GNP per capita (figure 1). In general, however, road maintenance expenditures depend not just on the size of the network but also on the nature of the terrain (higher costs in mountainous countries such as Bolivia and Nepal), the severity of the weather (higher costs in cold or wet climates such as Norway or Bangladesh), and construction standards. Insufficient or uncertain budgetary allocations undermine orderly planning and execution of road maintenance, resulting in road deterioration, which significantly increases transport costs. This article addresses the fiscal and institutional means for avoiding this outcome.

Deferring road maintenance increases not only total costs but also the present value of the future cost stream at any reasonable rate of discount. If roads deteriorate to the point at which they need reconstruction, restoring them to the original level of service costs three to five times more than timely and effective maintenance (Harral and Faiz 1988). Gyamfi (1992) estimates that the ratio of restoration costs to maintenance costs lies between 1.0 and 2.5 in Chile and Costa Rica. In addition, for every dollar that road agencies withhold by underfunding maintenance, road users must spend about three dollars in additional vehicle operating costs. Therefore, the economy as a whole (although not necessarily the road agency itself) benefits from timely road maintenance at any reasonable discount rate.

Figure 1. *Road Maintenance Expenditure Patterns in Selected Countries: Annual Averages for the Period 1988–95*

Source: Table A-1.

The World Bank's Operations Evaluation Department (OED) database supports e assessment that road maintenance is a highly productive expenditure. The OED tabase covers 341 road projects evaluated between 1961 and 1988. The average timated economic internal rate of return for this class of projects was 38.6 percent. his rate compares with an average of 26 percent for all transport projects and 21 rcent for all World Bank investment projects (Heggie and Vickers 1998).

In the 1960s and 1970s many countries in Africa, Asia, and Latin America estab-shed road funds as an extrabudgetary arrangement. Earmarking of a stream of rev-nues (often from fuel taxes) was introduced for use by the road department or gency, with the intention of insulating maintenance expenditures from erratic and equently delayed general budget procedures and reviews (Johansen 1989). (For a efinition of earmarking and other terms used throughout the text, see the glossary.) Then, the World Bank requested that these funds be set up to protect its invest-ents from economically inefficient asset deterioration. Some recently established oad funds in Eastern Europe (in Russia and Georgia) are still based on this model.

From a macroeconomic viewpoint, analysts have criticized earmarking because it nfringes on the policymaking powers of state executives and legislatures and reduces he leverage and flexibility of central governments (Deran 1965). Therefore the World Bank and the International Monetary Fund (IMF) have discouraged first-generation road funds in practice on general macroeconomic grounds (World Bank 1986). Even within the transport sector, some analysts have argued that the preferential access of road funds to lucrative revenue sources, such as gasoline taxes, hinders development f a more balanced, multimodal transport system.

The performance of state road agencies continues to be a matter of concern. There-fore a new generation of road funds is being established in the 1990s as part of an agenda to commercialize the road sector. This process runs parallel with (and prima-rily for the same reasons as) the privatization of state enterprises. Second-generation road funds are funded by levies or surcharges designated as user charges and identi-fied separately from general taxation. Revenues are paid directly into a fund man-aged by a road board whose membership is chosen to represent users. The road board determines both the level of charges and the allocation of expenditures. Second-generation road funds have been established in several African countries and in countries as diverse as El Salvador, Guatemala, Jordan, Lebanon, and Pakistan (Heggie 1995).

Off-budget financing, or earmarking, is not universally necessary or effective. Many European countries with good governance maintain road systems through traditional within-sector budgetary arrangements. Poor governance or governments' lack of self-discipline may make it impossible to maintain roads even with the existence of a statutory road fund. Governments may not be able to guarantee the security of as-signed revenue streams or the designated allocation of expenditures (Potter 1997a). Most countries fall between the extremes of good and bad governance. Assigning

responsibility for tax revenues and expenditures to a representative road board t
manage a specific road fund (similar to specialized taxing districts) may make a re
difference.

The next section sets out the analytical framework for assessing the desirability o
road funds, which has much in common with the framework for fiscal decentraliza-
tion. An important caveat is in order. Selective establishment of a road fund is justi-
fied only where quasi-commercial forms of organization might improve the alloca-
tion of resources through the implementation of a system of direct charges for road
infrastructure use and a system of fund governance by users. Road maintenance does
not inherently have a higher priority than other uses of public resources; thus even
though the discussion focuses on road funds, the general principles developed in th
article apply equally to other sectors where quasi user charges and user governance
can be implemented.

Analytical Framework for Considering Road Funds

The establishment of a road fund may affect the efficient working of the econom
through three main channels. The first channel, fiscal control, influences the effi-
ciency with which resources are collected and allocated among activities to maxim
total community welfare. The second channel, *management incentives*, partly deter-
mines the efficiency with which the agents of production use the resources alloca
to them. The third channel, rent-seeking behavior, can adversely affect both f
control and management incentives. Rent-seeking behavior occurs when individ
or agencies attempt to secure their own specific advantage at society's expense
1988). The relative importance of and balance among these channels critically
fects the assessment of the efficacy of road funds.

Fiscal Control and Allocational Efficiency

Public finance economists argue against earmarking because the optimal charge
road users would be unlikely to generate the revenue stream required to finance
optimal stream of road maintenance expenditures in terms of scale or timing. I
given time period, any earmarked stream of revenue is likely to generate either in-
ficient or excessive funding. Too little funding causes the road authority to re
continued recourse to the general budget. Too much funding creates the pote
for financing lower-priority expenditures. In the former case, ensuring independ
from poor budgetary processes and allocations requires rate-setting capabilities
the latter case, avoidance of wasteful allocations requires public accountability through
strict monitoring and auditing mechanisms. Often, the earmarking mechanism lacks
these features.

A related argument against earmarking arises from the fact that budgets in developing countries are often very fragmented. The development budget is typically a set of separate budgets, ringfenced by sponsoring donors. The recurrent budget is also fragmented by the prior call of debt servicing and other statutory expenditures and by the large share of wage expenditures that are difficult to cut. The proliferation of earmarked funds by sector or objective would make it difficult to rationalize the allocation of resources as expenditure priorities change. In short, a problem of macroeconomic control arises if extensive earmarking eliminates fiscal flexibility. This situation occurred in Colombia in the 1980s (Premchand 1983) and in Ecuador in the late 1990s.

The fiscal argument against earmarking in general and against road funds in particular is based on the assumption that general-purpose governments are better informed than special-purpose governments about the costs and benefits of alternative expenditure possibilities. The argument also assumes that general-purpose governments are committed to optimizing social welfare; that is, they are informed and benevolent. These assumptions may be true in some cases, where governance is representative and transparent. If the government system lacks these qualities, however, one of the main arguments against earmarking fails.

A related point is that current political pressures or the electoral cycle may result in myopic decisions. Vehicle operating costs, which do not enter into the road agency accounts, make up a high proportion of total transport costs (75–90 percent); these costs increase progressively as road conditions deteriorate (Harral and Faiz 1988). Unfortunately, road deterioration reveals its symptoms late. Expenditures on timely maintenance do not yield such obvious improvements in system performance as do expenditures on new investment. Yet long-term investment funded at the expense of optimum maintenance actually leads to a long-term decline in total available road system quantity and quality. The introduction of explicit road user charges, directed to a road fund in lieu of allocations from the general revenue budget, would contribute to allocative efficiency.

The introduction of explicit road user charges, however, would not automatically eliminate the need to address tradeoffs. In the absence of complete independence between specific road user charges and general taxes, securing funding for roads would entail an opportunity cost in other sectors. For example, in developing countries with low taxable capacity, fuel taxes represent a fairly secure tax source, accounting for 7 to 30 percent of total tax revenues and 1 to 3.5 percent of gross domestic product, or GDP (Gupta and Mahler 1995). The loss of control over this source of revenue may particularly damage the central government's economic management abilities. Introducing an indirect road user charge, in the form of a surcharge on fuel taxes, would limit the government's ability to increase taxes on fuel for general tax purposes. This limitation could lead to increased instability in the use of remaining tax revenue for social expenditures, such as health and education.

The independence of general taxing capacity from the level of road user charges i[illegible] likely to be greatest when a group of beneficiaries is well defined and the payment o[illegible] user charges is directly linked to the receipt of services. The benefit rationale for ear[illegible] marking aims to reveal taxpayer preferences for public services and to send a demand signal to the public sector about how much of the public service to supply (Bird 199[illegible]). With an efficient and fair charging mechanism, no one receives a service without pay- ing for it or pays without receiving the service. Second-generation road funds attemp[illegible] to honor this principle by generating the bulk of revenues through vehicle license fees, axle loading or distance fees, and, in a few cases, toll revenues. To a lesser extent, the[illegible] generate revenue through the separation of the pure tax element from the [illegible] explicit user charge element of public revenues collected from fuel taxes.

Management Incentives and Operational Efficiency

The life of a highway investment and the benefits accruing from it depend on the maintenance of the facility. Most appraisals assume optimal maintenance, although they may not explicitly address what this implies. Failure to provide the requir[illegible] maintenance effort means that the return on the initial investment will be lowe[illegible] normal budgetary practices do not provide the necessary funding for optimal m[illegible] tenance, then project designers and evaluators should reduce the likely benefit str[illegible] (and therefore the expected rate of return) or introduce complementary institution[illegible] mechanisms to ensure appropriate maintenance practices. In the first case, fe[illegible] investment projects would meet the criteria for selection. In the second case, est[illegible] lishment of a road fund to ensure funding for road maintenance from road us[illegible] charges (quasi prices) may be the logical corollary of accepting projects with attra[illegible] tive rates of return, but in contexts where budget practices are poor

The introduction of road user charges payable directly to a road fund can imp[illegible] managerial incentives by increasing autonomy from unwarranted political inter[illegible] ence. In many countries, wrangling over the budget delays its approval and disb[illegible] ment. Studies in Latin America show that uncertainty or untimely availabili[illegible] funding to maintain regular work schedules and to buy fuel and supplies expl[illegible] part the low equipment utilization rates and low number of kilometers maint[illegible] per employee (Gyamfi 1992). Even if the total level of road funding is open to [illegible] petition from other demands, a road fund may enable the executing agency to [illegible] form more efficiently by guaranteeing the availability of a secure core of fundin[illegible] Ghana the establishment of a road fund has substantially reduced the proble[illegible] disruption of the planning and execution of maintenance work. These problems [illegible] caused by delays in approving the budget or releasing the budget allocations and [illegible] of synchronization between the budget year (the calendar year) and the constructio[illegible] season (September to May). These delays necessitated the awarding of small [illegible] tinuation contracts to contractors who already had commitments with the autho[illegible]

tration. The establishment of road funds has eliminated payment delays, given a significant boost to contractor cash flow, and reduced unit costs by 15 to 20 percent (Pankaj 1989).

The guarantee of a core of finance may also allow road agencies to extend and improve contracting arrangements with the private sector. The same studies for countries in Latin America and the Caribbean suggest that maintenance by force account (that is, staff on government payrolls) is little more than half as efficient as maintenance that is contracted out to the private sector. In Ghana the greater certainty of funding associated with earmarking allowed the introduction of effective competitive bidding. In general, more reliable financial arrangements lead to better use of resources.

Operational efficiency may also increase if users willingly pay for maintenance because the road authority channels payments more directly to the provision of a service of value to the users. (The availability of these additional resources, which might not be forthcoming otherwise, can also improve the government's ability to manage macroeconomic imbalances.) Some countries, including many in Sub-Saharan Africa, have experienced a severe crisis in the maintenance of their main road networks. Heavy users, such as truckers and other operators of commercial vehicles, have demonstrated a willingness to levy an additional charge on their "own" use of fuel to finance a road fund with responsibility for maintaining a core network. There is no mystery to this behavior. Users more than recoup the surcharge if it is dedicated to fund better road maintenance, which in turn reduces vehicle operating costs. For example, Heggie (1995) estimates that vehicle operating costs decrease by two to four dollars for every additional dollar spent on road maintenance.

Rent-Seeking Behavior and the Distribution of Welfare

At the heart of the problem of traditional road funds was the failure of the associated earmarking arrangements to address incentive and governance issues. Unlike marketable commodities, including deregulated rail and airline services, the typical traditional road fund had no link between the tax rate (or the amount of taxes earmarked) and spending priorities (in light of the level of road use). Road fund managers had incentives to maximize their discretionary expenditures (including investment in low-priority roads or ancillary activities) rather than to optimize the level of road maintenance. The combination of public scrutiny and periodic monitoring by a competent central bureaucracy may provide some defense against this problem in industrial countries, but that combination is less likely in developing countries with less developed institutional capabilities.

Public choice theorists express skepticism even about institutional capabilities in industrial countries, including the political process that translates citizens' preferences into public action. Essentially they believe that diverse citizens' preferences do not permit aggregation into a well defined community preference function. They

also believe that monitoring costs and informational asymmetries may enable public officials (regardless of whether they respond to organized pressure groups) to project their personal interests onto their function of allocating resources. And they believe that budget choices do not depend solely on the inherent costs and benefits of services but also on the ability of one set of taxpayers to transfer the costs of programs that benefit them to others.

Where individual preferences for public goods differ, separate earmarked funds could potentially increase general welfare if the payments to those funds reflect individuals' relative marginal utilities for different public goods (Johansen 1963). Despite the ingenuity devoted to designing ways of getting consumers of public goods to reveal their marginal utilities truthfully, this analysis remains difficult to apply practically. Quasi prices, or user charges, may have welfare advantages because they can be levied approximately in proportion to the demonstrated benefit of consumption.

Public choice theorists have pointed out a fundamental flaw in general fund budgeting. That is, heavy consumers of a service that is financed through general taxes would benefit from lobbying for larger expenditures on that service (thereby transferring welfare to themselves). At the same time, nonconsumers would argue for lower expenditures. The outcome depends on the respective political power of the parties, rather than the aggregate value attached to each individual service. The road authority can eliminate this bias by setting prices, such as tolls and vehicle duties, for the beneficiaries of a specific service. Using a fuel surcharge as a quasi price for road use (with appropriate corrections for agricultural vehicles and for fuel not used for road vehicles) is analagous to establishing a special taxing district. These districts are common in the provision of some facilities, such as water, and could be consistent with the government's pursuit of redistribution objectives through its policies on general taxation and the allocation of merit goods.

The argument for earmarking as a way of separating allocation and distribution issues may also be applied spatially. The government could implement a program of regional financing for services consumed regionally. This program would avoid overprovision in some regions at the expense of others as the regions compete to maximize their share of the national budget. Of course, the road authority may justify some regional disparities in provision, particularly of road investment, on both efficiency and equity grounds. The government would require operational criteria for the spatial allocation of resources, as have recently been developed for second-generation road funds, regardless of whether the funds are earmarked.

Empirical Evidence

The influences discussed above do not militate in the same direction and may demonstrate variable quantitative significance in different circumstances. Therefore the

evaluation of second-generation road funds cannot be resolved solely by reference to a priori theorizing but instead must be approached on a case-by-case basis in light of national circumstances. Several issues affect the balance between the micro- and macroeconomic considerations, including the demonstrated effectiveness of road funds in meeting their public service objectives, the magnitude of the economic cost of undermaintenance, the extent of the backlog of road maintenance work, the demonstrated capability of existing fiscal arrangements to finance efficient maintenance, and the degree of current fiscal dependence on taxation of road users.

Experience with the Operation of Road Funds

The empirical evidence suggests a positive but statistically weak ability of traditional forms of earmarking to alter resource allocation significantly (Eklund 1967). Wherever the government retains control over the level of the user charges or over the allocation of complementary funds, the total level of funding may be just as vulnerable with a road fund as without one. For example, the earmarked funds for the Colombian National Road Fund grew at the same rate as GDP in 1979–87, but road expenditures per kilometer fell by 25 percent (Dick 1989). The fall in expenditures occurred in part because funds were preempted to cover increased debt service payments.

McCleary (1991) reviews the experience in the Central African Republic, Colombia, Ghana, Mali, and Zaire (now the Democratic Republic of Congo) in the late 1980s. He concludes that the road fund was successful only in Ghana, where the government was strongly committed to increased expenditure on roads in any case. In general, the governments found it difficult to set appropriate levels for earmarked taxes. Allocations tended to continue to depend on the general budget situation, and adequacy of overall road fund resources provided no assurance of an appropriate balance among maintenance, rehabilitation, and new investment. A more recent review of 10 African road funds confirms many of these failings (box 1). It also emphasizes the need to examine more closely why the funds failed. In contrast, second-generation road funds overcame many traditional defects when they had clear objectives, an independent revenue source, and efficient management and accounting arrangements (Balcerac de Richecour and Heggie 1995). Some newer funds (for example, in Ghana and Zambia) ensured the automaticity of payment by separating road levies from general taxes.

Experience with Existing Fiscal Arrangements

Analyses of the views of country road agencies and task managers in nine countries shows the ratio of actual maintenance expenditures to those considered necessary to maintain constant road conditions. The ratio varies between 29 percent for Jordan and 59 percent for Argentina, with most countries falling in the 50–70 percent range

Box 1. *Major Problems with Conventional Road Funds, with Illustrative Examples*

Continued inadequate maintenance of the road network

Insufficient revenue base. The revenue base of the road fund is large enough to finance only part of qualifying expenditures, with the balance financed through the government's general budget [illegible] road agencies obtain some money from the road fund, they may find it even more difficult [illegible] funds through the normal budgetary process.

- *Ghana.* The road fund's original revenue base could finance only 60 percent of periodic [illegible] tenance.
- *Zaire.* In 1985 a higher fuel tax increased the revenue base of the existing road fund, but [illegible] insufficient to cover all of the qualifying expenditures. In 1986 the new budget [illegible] dropped to zero, resulting in a drop in total road funding. The situation was repeated [illegible]

Funds raided by the general budget. The revenues are collected by the customs and excise [illegible] and channeled through the ministry of finance, which almost invariably withholds some [illegible] revenues.

- *Sierra Leone.* In February 1990 the central customs authority reduced the amount paid [illegible] road fund and in June 1991 suspended payments altogether, even though the road [illegible] tinued to pay the same fuel levy.
- Similar situations occurred in *Ghana, Mozambique,* and *Tanzania.*

Poor governance, limited operational efficiency, and misappropriation of funds. Without [illegible] dit procedures, the government cannot track how monies paid into the road fund are spent [illegible] the government ensure that only approved expenditures are covered and that these expenditures [illegible] required specifications.

- *Mozambique.* Audit reports for 1993 and 1994 could not verify the flow of monies [illegible] Petromac (the oil company that imports fuels), the Ministry of Finance, and the road [illegible] could the audit determine whether the road fund had received all the revenue due [illegible] levies on gasoline and diesel fuel.
- *Rwanda.* The 1991 audit report was unable to certify accounts because of [illegible] financial information and a lack of specific information on the revenue side.
- *Sierra Leone.* The audit for September 1989 to December 1991 showed that about [illegible] was used to purchase vehicles that were never delivered.
- *Tanzania.* The audit report for fiscal year 1992/93 showed that about [illegible] ments was made without any supporting vouchers or other documents.
- *Zambia.* $760,000 of the revenue collected for roads between May 1993 and [illegible] never reached the road fund. The audit also showed that about $500,000 was paid [illegible] that were never supplied, about $6 million was paid to contractors without any [illegible] to show that work had been done, and about $70,000 was paid for the purchase of [illegible] were never delivered.

Misallocation of resources

Diverting funds to low-priority and nonrelated activities to avoid showing a surplus. When [illegible] finances roads managed by different agencies without any transparent and equitable [illegible] allocating revenues among them, allocations are often subject to political whim rather than [illegible] criteria. A weak or nonexistent road fund board and ambiguous legislation add to the problem.

- *Central African Republic.* The 1993–94 audit report showed that the central government [illegible] an irregular loan of $340 million from the road fund to pay civil service salaries.

Sierra Leone. From September 1989 to December 1991, there were no guidelines regarding use of funds. The presumption was that the funds were intended for road maintenance. Substantial sums were spent on refurbishing offices, purchasing 1,800 yards of carpet, and carrying out repairs at the State House and the Parliament building.

Tanzania. The audit report for fiscal year 1992/93 found that the road fund was used to finance recurrent expense items not covered by the directives issued by the Ministry of Finance, including leave payments, ferry operators, electric bills, and gratuities. Of the funds allocated for urban and rural roads, 75 percent went only to urban areas because they were better able to prepare plans — not because their needs were inherently greater.

Zambia. Between May 1973 and November 1993, $4.[illegible] million was released through provincial accounting control units for road rehabilitation. No guidelines for disbursing funds were provided, no expenditure returns were submitted, and no expenditure returns were requested before further funds were released. About $600,000 was paid for items that should have been covered through normal budgetary provisions.

...venue flow into the road fund. A persistent buildup of a surplus that is not accessible by the ...al budget can severely limit the financial resources of the central government in a fiscal crisis.

Colombia. In 1986 more than 35 percent of central government tax revenue was earmarked, with the road fund alone accounting for nearly [illegible] percent.

...governance because the governing agency is captured by those who shift allocative priorities.

South Africa. Its initial road fund board, which primarily consisted of provincial representatives, found it difficult to get the members to act in the national interest.

...: Johnson (1989), Heggie (1995) [illegible]

...98) Table 1 shows even more pessimistic conclusions from assessments of maintenance provisions based on the World Bank's Highway Design and Maintenance Standards model for World Bank projects.

...klog of Road Maintenance Work

...study of road deterioration in developing countries, Harral and Faiz (1988) ...mate the annual maintenance expenditures required to prevent deterioration. For ...[illegible] expenditures varied from 0.2 percent of GDP, on average, for countries in ... Asia and the Pacific to 1.0 percent for countries in West Africa. Harral and Faiz ...ate that the backlog of maintenance work varied from 1.6 percent of GDP in ... Asia and the Pacific to 3.5 percent in South Asia. Although these calculations ...not been updated recently, anecdotal evidence suggests that the pattern persists. ...se estimates in table 1 do not suggest that lack of maintenance in the road sector ...worse than the underfunding of maintenance expenditures in the provision of ...er public goods and services. And they do not suggest that the recommended ...ls of expenditure were socially optimal in economic terms. The estimates in table 1 ...w only the extent to which funding was insufficient to maintain constant service ...els for an economically critical asset.

nstruments, including a surcharge on fuel taxes. In the longer term, a commercial oad agency would replace the road fund. Tolls charged on a fee-for-service basis or harges more directly related to the costs imposed by road use (for example, charges or trucks based on axle weight and distance traveled) would replace the fuel surharge. New Zealand is already moving in this direction.

ption 2: A Reformed Budget Process

n the second main option, the government would rebuild the capability of, and onfidence in, its budgetary processes. In some cases, commercialization is not potically or practically feasible, and the public is unwilling to pay additional general xes. Earmarking some special-purpose taxes and creating a road board–cum–road und may provide the most practical interim means for generating additional funds or a priority economic activity and rebuilding public confidence in government. ven weak governments may accept some transient impediments to raiding activities with high benefits but low profiles, thereby changing the balance of expediency. n the longer term, the government would phase out the road fund and return all evenue and expenditure responsibilities to the budget. Countries in the European nion commonly rely on the budget rather than on road funds.

Interim Arrangements

he decision whether to establish a road fund as an interim step in option 1 or ption 2 is a complex issue and requires case-by-case analysis (box 2). Three principles guide the decisionmaker about whether to introduce (or retain) a road fund:

- Will it improve resource allocation—for example, by ensuring better funding of activities with economically high returns but politically low profiles (bearing in mind that other services such as primary schooling and basic health clinics may also fall in the same category)?
- Will it improve operational efficiency—for example, through the introduction of better incentives for managing resources?
- Will it reduce rent seeking—for example, by strengthening the link between benefits and payments?

Expenditure and Revenue Assignments and Governance

Assuming that conditions justify the creation or continuation of a road fund, its charter should explicitly address three broad issues. First, what road expenditure line items should the road fund protect, and what is the primary purpose of the fund? Second, what revenue streams or revenue authority should finance the chosen expenditure items

Box 2. *Conditions for Introducing a Road Fund*

Introduction of road funds may be justified if all the following conditions apply.	*Potential indicators*
Maintenance is poor because of: • *Insufficient funds.* Poor setting of budget priorities with bias in favor of new investments, often donor driven. • *Unreliable timing of funds.* Poor budgetary processes that cannot ensure credible commitments, disbursements, or both. • *Inefficient implementation of works.* Absence of incentives to use resources efficiently in the agency.	• Asset condition of core network predicted to decline over the next 10 year period (increasing percentage in poor condition by [illegible]) • User costs predicted to increase over [illegible] (increasing ratio of vehicle operating [illegible] vehicle/year by vehicle class) • Substantial maintenance forgone (economic rate of return greater than 20 [illegible]) • Net present value of near optimal [illegible] more than 1.5 times higher than the current program • Total costs per mile of road maintained on the core expected to drop by [illegible] throughout the life of the fund [illegible] current expenditures or future [illegible]
There is political commitment to increase maintenance expenditures on roads	• Cabinet level commitment (for example, gazetting) to increase road [illegible] expenditures • Cabinet level commitment to [illegible] user charges or surcharges [illegible] to generate funds • Cabinet level commitment [illegible] parallel funding
There is a political commitment to establish long term reliable mechanisms for improved allocation and accountability for the core network	• Principles accepted for [illegible] decisions • Representatives of key [illegible] on the road board • Economic criteria accepted as [illegible] priorities

Note: The size of the core network will vary over time as unused and lightly used [illegible] and emerging heavily used routes are included.

in the road fund? And third, how will the proposed institutional structure [illegible] sponsible governance and appropriate incentives to reconcile the conflict [illegible] and macroeconomic management objectives? In particular, what features [illegible] that unrepresentative or unaccountable interests will not capture the funds

Expenditure Assignments

Should the government or the road authority decide about the allocation of [illegible] between investment and maintenance, among regions or road types, and [illegible] administration and implementation activities?

ALLOCATION OF RESOURCES BETWEEN INVESTMENT AND MAINTENANCE. The most commonly identified problem, systematic bias against maintenance, occurs in fiscal regimes that fund both investment and maintenance through the same channels (with or without road funds). The regime tends to favor investment because large schemes are politically attractive and have easier access to external financing. In addition the benefits of investment are more apparent ex ante than the benefits of maintenance. As a result, decisionmakers behave as if their discount rate were greater than the technical rate. This problem commonly occurs in developing countries as well as in transition economies, such as in Eastern Europe.

The creation of a road fund with both investment and maintenance responsibilities does not automatically ensure against such a systematic bias. In Mali the road fund provided the national counterpart funding for foreign lending for road construction and financed the servicing of debt on earlier investments. These became priority calls on the road fund, and maintenance suffered accordingly. This suggests that road funds be dedicated to maintenance, thereby providing a needed counterbalance to a systematic bias.

Experience provides less clear guidance on the treatment of investment. Several countries (including Japan, the Republic of Korea, South Africa, and the United States) introduced road funds to facilitate crash investment programs. The governments considered the investment programs too large for the general budget, thus justifying special treatment, including extra special purpose taxation. But such arrangements can create a temptation to misallocate funds to lower-priority investments if they continue to generate large amounts of revenue after the real need that prompted their creation has been satisfied. This problem, combined with the systematic bias in favor of investment, would appear to provide sufficient reason to exclude investment from any ringfenced allocation in an interim arrangement.

Linking closely related investment and maintenance requirements, however, is appropriate over the long run. Freed from the responsibility of funding the operating and maintenance consequences of their investment decisions, governments and [illegible] could continue to indulge in excessive road investments, leading to debt-[illegible] and maintenance funding burdens that, combined, are unsupportable. This [illegible] would argue in favor of a commercialized road agency or road fund, recoup-[illegible] the costs of investment and maintenance. For the road agency or road [illegible] influence investment expenditures, however, it would need a broad mem-[illegible] extended beyond direct road users (for example, to groups displaced or [illegible]enced by new road construction). It would be politically difficult to del-[illegible] right-of-way and eminent domain functions to an entity such as a road [illegible]

[illegible]ON OF RESOURCES AMONG REGIONS OR ROAD TYPES. Another reason for [illegible]ding investment is that road boards dominated by user representatives may not

allocate investment resources optimally beyond the narrow confines of road maintenance. Roads perform social and strategic, as well as economic functions, which may justify cross-subsidies. For example, governments keep some rural roads in existence and repair even though their users cannot pay sufficient sums to maintain them. Ecological or aesthetic reasons in environmentally sensitive areas could require additional expenditures on roads beyond the amounts users would willingly pay. The same is true for strategic functions. There are several ways to approach this issue, but all of them introduce complications in the governance of road funds.

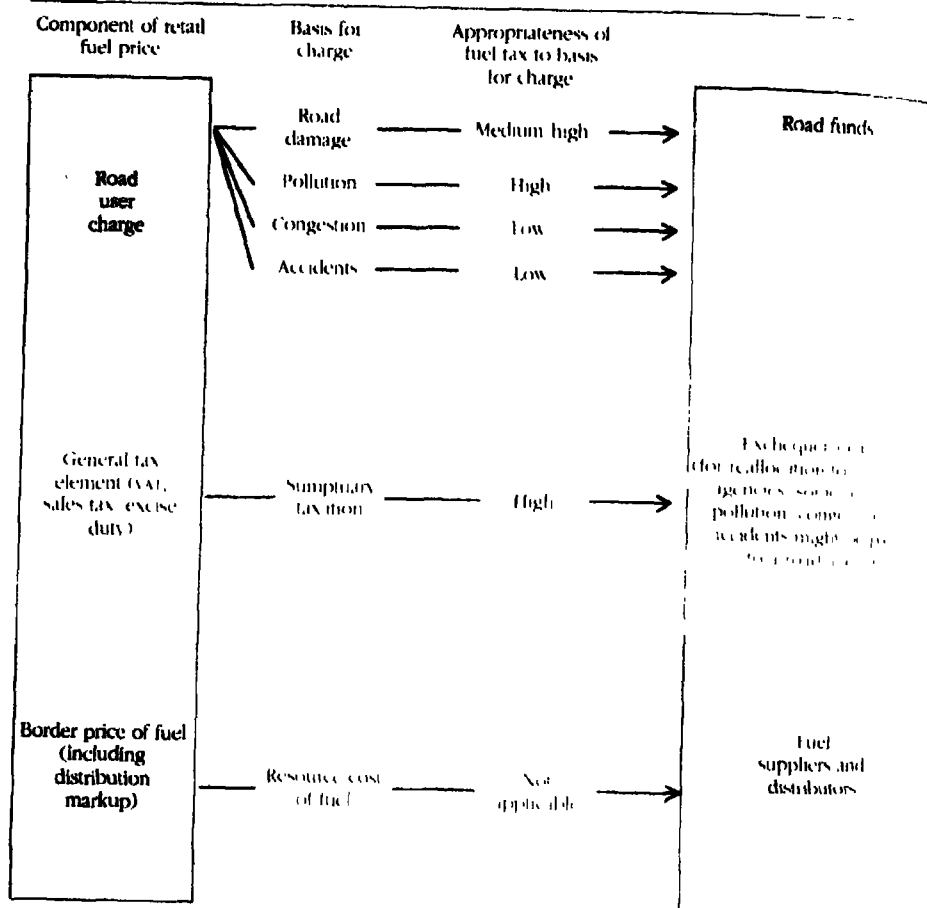
First, the road authority could introduce a system equivalent to the public service obligations of transit operators. Such a system would use a contracted payment from the state to compensate the commercial road fund for meeting an explicit public obligation. This arrangement reduces the neatness of the separation between the road fund administration and the political process and reintroduces substantial scope for negotiations about financial responsibility.

Second, the government could expand the road board to include representatives of noncommercial, environmental, and local interests in the management of the road fund. The expanded road board would address issues of acquisition of rights of way, resettlement, and other problems associated with the expansion of the road network. This solution, too, would reintroduce an element of politicization.

Third, the government could create multiple agencies, each concentrating on a more restricted set of roads with representative management. This arrangement, however, could restrict legitimate transfers across regions or types of roads. Box 3 summarizes the issues created by multiple road funds.

Box 3. *Single or Multiple Road Funds*

Single road fund	*Multiple road funds*
National	By type of road (rural, interurban, urban) By administrative unit (district, province)
Benefits	
• Potential to cross-subsidize between high-use and low-use links, particularly if the fuel surcharge applies to all road users irrespective of the specific links they use	• Potential for charge discrimination [illegible] on relatively distinct user segments
Issues	
• How to design a transparent spatial or functional allocative formula, for example, by population, length of road network, type of network, priority.	• Whether each fund should [illegible] revenue base, allocative criteria [illegible] governance structure
• How to ensure the representativeness of the board, especially for the purpose of setting priorities and allocating resources.	• How to deal with components [illegible] national roads that are part of the [illegible] network

Figure 2. *Gas or Diesel Fuel as a Base for Road Maintenance Charges*

Source: Authors' calculations.

Fuel use, however, is not as good a proxy for road damage by trucks, which account for the bulk of road damage. The impact of a heavy truck is 10,000 times as great as that of a medium auto (Newbery 1987).[1] It also varies nonlinearly with axle weight. Truck use would require a very high surcharge on diesel fuel, thereby creating problems for the use of diesel outside the transport sector (as in power generation or agriculture). Otherwise, the charge is too low to collect the necessary revenue or to affect truck use.

A surcharge on fuel may be appropriate for other externalities, such as air pollution, but not for other externalities such as congestion (which varies by time and location) or safety.[2] Given the multiple claims on the surcharge, the road authority need not direct all the revenue generated by the surcharge to a road administration

a road fund, or a commercialized road agency. Some of the road user charges will have to be transferred to other budgets (see figure 2).

If a fuel charge is used, the road authority should base the design on a transparent and supportable predetermined formula that is directly linked to the externalities or services consumed. The introduction of a fuel tax surcharge as a quasi price for road use should be approached particularly cautiously if the following circumstances apply:

- Taxation on fuel is a high proportion of total tax revenue (say, more than 10 percent), and any surcharge is likely to detract from general taxing capability. This situation occurs in many small developing countries. For example, the proportion is 19 percent in Costa Rica, 16 percent in Guatemala, and 13 percent in Nepal.
- The general fiscal situation is weak and many sectors are making similar claims for special fiscal treatment (as in many countries with internal deficits in excess of 5 percent of GDP or tax-to-GDP ratios of less than 15 percent).
- Diesel fuel is used extensively in power generation and in agriculture (as in many countries in South and East Asia).

The road authority should anticipate and avoid the danger of resistance to supplementary funding from the general budget once the road fund has been set up. Road maintenance may still require funding from the general budget. Insofar as the efficiency arguments in favor of a partially protected budget relate to the ability to plan [illegible] core maintenance work program, the benefits of a road fund do not [illegible] on complete separation of the road budget from the rest of the budgetary [illegible]. Funding for links outside the core network or to cope with unforeseen circumstances (for example, natural disasters) would still require appropriations from [illegible] budget.

[illegible]

[illegible] of the argument in favor of second generation road funds is the separation of user charges from general taxes and improved arrangements for governance [illegible] funds. Many traditional road funds failed because funds accumulated or [illegible] allocated. Better governance is essential to ensure that budget constraints [illegible] and that expenditure decisions are responsive to users. The following four [illegible] components should therefore be included in the package.

The package should have a strong legal basis. Road funds should be established [illegible] law to ensure clear terms of reference and some minimum protection from arbitrary political interference. The legal instrument not only needs to guarantee the [illegible] of funds but also must ensure that the funds are automatically channeled to

the executing agency.[3] Where the setting of user charges and taxes overlaps betwee
the road board and the government, the instrument should establish clear proce
dures enabling the executing agency to have the greatest possible security of a bas
level of funding and the greatest possible notice of changes in that base. It is unrea
sonable, however, to expect legal commitment to any particular level of tax or ta
yield.

2. The package should have an independent executive authority. The need fo
efficient maintenance suggests that the executing agency should be accountable in
clear and transparent framework. It should have the primary role of formulatin
maintenance policy, marshaling and allocating funds, and securing effective imple
mentation. Where these functions are clearly stated and well publicized as the re
sponsibilities of a quasi-independent executive, as in Brazil and Chile, they establish
proper management incentives and facilitate effective performance

3. The package should have a third-party monitoring system. Given the problem
of securing representative governance, a monopoly supplier of road services (the op
erator of the road fund or commercialized road authority) should be subject to regu
latory supervision. The fund or authority should be required to inform the public
and supervisory authorities of its activities; its accounts should be externally audit
and periodically reviewed by an independent review body; and, if corporatized
net income should be subject to corporate income tax.

4. The package should delineate administrative competence and proper crite
for expenditure. Assigning expenditure responsibilities to the road fund does not pe
se ensure efficient allocation within the ringfenced area. As necessary conditions
assigning revenues, the executing agency should have well-established procedures
allocating funds efficiently and the necessary administrative competence to admi
ter the allocation and to monitor and report on performance. Mozambique and Ne
Zealand have developed effective procedures for allocation of funds (Heggie a
Vickers 1998).

Where the road fund is established as a temporary arrangement until general b
getary procedures are improved, termination criteria and a sunset clause must
place to determine what should occur when effective budgetary procedures are
to have been reestablished. Box 4 summarizes suggested criteria for the termin
of earmarking taxes to a road fund based on an independent review.

Where the road fund is established as a stepping-stone to a commercial
because the government wants to disengage from the direct production of go
and services, a review should be commissioned. If the review determines that
performance criteria set out in box 2 have been satisfied, it should then form
the road authority as a commercialized public utility with user charges acc
directly to it rather than passing through the government treasury. Subsequently
the road authority should be subject to the same general form of public scrutiny
other privatized monopolies. The response to an unfavorable review should be

Box 4. *Sunset Provisions for Road Funds*

Closure of road funds may be justified if the following apply.	*Potential indicators*
Continued inadequate maintenance of the road network because of: • Insufficient revenue base that is not topped up as needed by the general budget. • Raiding of funds by the general budget. • Poor governance due to incentives that limit the operational efficiency of the fund.	• Ratio of revenue flow to the fund to the required (nearly optimal) maintenance expenditure flow less than 0.7. • Ratio of actual expenditures to required maintenance expenditures less than that of the revenue flow to required expenditure, and the actual expenditure flow also less than that before institution of the fund. • Increase over time of predicted real vehicle operating costs for the roads maintained. • Actual maintenance cost per kilometer of road greater than comparable benchmark. • Percentages of roads in fair and poor condition increasing.
Misallocation of resources • Wasteful spending to avoid showing a surplus • As a result of high revenue flow into the fund a persistent buildup of a surplus that is not accessible by the general budget	• Ratio of administrative expenditure to total expenditure greater than 20 percent. • Actual expenditures greater than the estimated required maintenance expenditures.
Poor governance due to capture that shifts allocative priorities	• Inadequate representation of stakeholder groups in board management structure. • Inadequate public reporting of plans, expenditures, and road conditions. • Maintenance expenditures on links not in proportion to the corresponding economic rates of return from maintaining the links.

in most countries, a matter for political and legislative rather than simple administrative action.

Conclusions and Recommendations

Road funds are a continuing subject of controversy. Highway specialists regard them as a boon to facilitating the provision and maintenance of a highly productive asset by means entirely consistent with the general shift away from direct government production of goods and services. Macroeconomists and public finance specialists have tended to regard them as a bane because they reduce fiscal flexibility, do not

adequately address problems associated with the provision of public goods or th
internalization of externalities, and are often not well managed.

This article has argued that the issue is not one to be resolved on general principle
but on a case-by-case basis through the analysis of likely micro- and macroeconomi
effects. In general, two radically different long-term options reconcile fiscal pru
dence with asset maintenance. The first is a fully commercially operated road agen
subject to the normal oversight of behavior accorded to privatized monopolies. Th
second is a reformed and well-functioning budget process. Neither option exists
present in most developing and transition economies. Thus, any recommendation
on the role and nature of road funds must be viewed as a provisional, case-specifi
intermediate step in the direction of one of the long-term solutions (Potter 199

This article lists a few indicators that can be used. In addition, the article sug
that the interim arrangement requires four conditions. First, the road fund's
penditure responsibilities should be limited to maintenance in order to corre
systematic bias against maintenance despite the link between investment and main
tenance. Second, the road fund's revenues should come only from direct charges on
road users, except in the case of fuel surcharges (which need to be separated fro
general taxes in agreement with the treasury). Where feasible, weight-distance charg
should be introduced for trucks. Third, the road fund should have professio
management, under the direction of a user representative board and subject to s
oversight and auditing arrangements by third parties. Fourth, the road fund requir
explicit transition arrangements as it moves toward a long-term solution

As recent experience in Africa shows, when there is a crisis and main road syste
fail, commercial interests can be mobilized to pay a surcharge on the existing fu
so long as the user charge is devoted to improving the quality of the road infrastruc
ture. Experience with second-generation road funds is very limited, however. It
mains to be seen whether road fund management will be sufficiently immune
kind of political interference that currently disturbs the flow of funds into road m
tenance. And road authorities need to overcome the immediate, most extreme p
lems of deficient maintenance.

Therefore, analysts and policymakers should monitor carefully the recently
duced road funds in Sub-Saharan Africa and elsewhere. Sector programs and in
ment projects establishing these funds should include appropriate and comparable
toring and evaluation components.[9] Monitoring and evaluation of second gen
road funds well help to determine the utility and applicability of such interim arra
ments and facilitate the amendment of initial designs on the basis of experie

Notes

Ken Gwilliam is a transport economist with the World Bank's Transport, Water, and Urban Devel
opment Department, and Zmarak Shalizi is a research manager with the Development

Table A-1 *(continued)*

Country	*GNP per capita (1995 US dollars)*	*Total government expenditure (% of GNP)*[a]	*Total road expenditure (% of total government expenditures)*[b]	*Road maintenance expenditure*	
				% of total	*% of GNP*
Turkey	2,832	20.28	3.26	34.63	0.23
Brazil	3,152	51.70	0.38	30.05	0.06
South Africa	3,195	45.24	1.73	32.37	0.25
Slovak Rep.	3,371			28.19	
Hungary	3,686	70.93	0.89	29.07	0.18
Korea, Rep.	7,629	16.68	12.00	9.70	0.19
Average[d]	2,160	31.80	7.22	21.03	[illegible]
Countries with per capita income $8,000 or greater					
Portugal	8,386	44.07	2.40	6.09	0.06
Ireland	13,429	61.98	2.10	34.60	0.45
Spain	13,771	50.61	2.1[illegible]	19.81	0.22
United Kingdom	19,205	52.26	1.56	40.82	0.3[illegible]
Australia	19,484	46.4[illegible]	2.85	29.68	[illegible]
Italy	20,853	54.93	2.60	33.66	[illegible]
Netherlands	22,479	69.45	0.70	57.20	[illegible]
France	24,209	54.41	1.04	11.11	0.06
Austria	24,42[illegible]	[illegible]	2.96	32.06	[illegible]
Finland	24,858	62.69	3.06	52.53	1.0[illegible]
Germany	26,543	55.80	1.79	11.51	[illegible]
United States	27,047	42.94	2.89	32.[illegible]5	[illegible]
Sweden	27,659	72.00	1.31	66.28	[illegible]
Denmark	28,236	[illegible]	1.16	52.59	[illegible]
Norway	31,067	62.3[illegible]	2.05	32.69	[illegible]
Japan	33,761	19.01	10.54	14.14	[illegible]
Switzerland	39,760	44.45	1.82	18.45	[illegible]
Average[d]	23,833	54.65	2.53	32.09	[illegible]

Not available

Note. Gross national product (GNP) per capita is based on the market exchange rate. [illegible] capita in terms of purchasing power parity shows that while values change for all the countries selected [illegible] in a change in internal rankings, there is no shift in countries between the two groups.

a. Total government expenditure includes all expenditures at the central, state, and local govern[illegible] Original data in local currencies are converted to dollars using the exchange rate data from the World [illegible] ment Indicators database of the World Bank.

b. Total road expenditure includes all capital and maintenance expenditures on roads at the [illegible] local government levels. Original data in SDR (Special Drawing Rights) are converted to dollars [illegible] change rate data from IMF, *International Financial Statistics Yearbook 199*[illegible]

c. Road maintenance expenditure includes all maintenance expenditures on roads at the [illegible] local government levels converted to dollars using the exchange rate data from IMF, *International* [illegible] *tics Yearbook 1997.*

d. The average for each variable is computed as the sample mean for the group of selected countries.

Source: World Bank (1997); IMF (1997), International Road Federation (1992, 1997).

Deran, Elizabeth Y. 1965. "Earmarking and Expenditures: A Survey and a New Test." *National [illegible] Journal* 18:354–61.

Dick, Malise C. 1989. "Earmarking of Transport Funds in Colombia." In Frida Johansen, ed.

Eklund, Per. 1967. "Earmarking of Taxes for Highways in Developing Countries." Working Paper, World Bank, Economics Department, Washington, D.C. Processed.

Gupta, Sanjeev, and Walter Mahler. 1995. "Taxation of Petroleum Products: Theory and [illegible] Evidence." *Energy Economy* 17(2):101–16.

Gyamfi, Peter. 1992. "Infrastructure Maintenance in LAC: The Costs of Neglect and Options [illegible] Improvement." Regional Studies Program, vol 4, report [illegible] World Bank, Latin America [illegible] Caribbean Technical Department, Washington, D.C. Processed.

Harral, Clell G., and Asif Faiz. 1988. *Road Deterioration in Developing Countries: [illegible] edies.* World Bank Policy Study. Washington, D.C.

Hau, Timothy D. 1992. "Economic Fundamentals of Road Pricing: A Diagrammatic [illegible]" Policy Research Working Paper Series 1070. World Bank, Infrastructure and [illegible] ment Department, Washington, D.C.

Heggie, Ian G. 1995. *Management and Financing of Roads: An Agenda for Reform.* World Bank [illegible] nical Paper 275. Africa Technical Series. Washington, D.C.

———. 1996. "Commercialization of Roads." Best Practice Discussion Paper. World Bank [illegible] ington, D.C. Processed.

Heggie, Ian G., and Rodrigo Archondo-Callao. 1996. "Road User Charges Model [illegible] Transportation, Water, and Urban Development Department. Washington, D.C.

Heggie, Ian G., and Piers Vickers. 1998. *Commercial Management and Financing [illegible]* Bank Technical Paper 409. Washington, D.C.

IMF (International Monetary Fund). 199[illegible] *Government Finance Statistics Yearbook.* [illegible]

International Road Federation. 1992, 199[illegible] *World Road Statistics.* Washington, D.C.

Johansen, Frida, ed. 1989. "Earmarking, Road Funds, and Toll Roads: A World Bank [illegible] Report INU 45. World Bank, Washington, D.C.

Johansen, Leif. 1963. "Some Notes on the Lindahl Theory of the Determination of [illegible] tures." *International Economic Review* (September) [illegible]

McCleary, William A. 1991. "Earmarking Government Revenues: Does It Work [illegible] *Research Observer* 6(1):88–107.

McCleary, William A., and Evamaria U. Tobon. 1990. "Earmarking Government Re[illegible] lombia." Policy Research Working Paper 425. World Bank, Country Economics [illegible] Washington, D.C. Processed.

Newbery, David M. 1987. "Road User Charges and the Taxation of Road Transp[illegible] Paper WP/87/5. International Monetary Fund, Washington, D.C.

Pankaj, Thampil. 1989. "Road Fund Experience in Ghana." In Frida Johansen, ed.

Paterson, William D. O. 1988. *Road Deterioration and Maintenance Effects: Models [illegible] Management.* Washington, D.C.: World Bank.

Potter, Barry. 1997a. "Dedicated Road Funds: A Preliminary View on a World Ba[illegible] Working Paper WP/97/7. International Monetary Fund, Washington, D.C.

———. 1997b. IMF Summary of the Meeting on Road Funds, memo, January. Washington [illegible]

Premchand, Arigapudi. 1983. *Government Budgeting and Expenditure Controls: Theory [illegible]* Washington, D.C.: International Monetary Fund.

Teja, Ranjit. S. 1988. "The Case for Earmarked Taxes: Theory and an Example." Working Paper 88/18. International Monetary Fund, Washington, D.C.

Between the State and the Market: Can Informal Insurance Patch the Safety Net?

Jonathan Morduch

Households in low-income countries deal with economic hardships through informal insurance arrangements between individuals and communities rather than through publicly managed programs or market-provided insurance schemes. Households may, for example, draw on savings, sell physical assets, rely on reciprocal gift exchanges, or diversify into alternative income-generating activities. These mechanisms can be highly effective in the right circumstances, but most recent studies show that informal insurance arrangements are often weak. Poor households, in particular, have substantial difficulties coping with local, idiosyncratic risks. Public policy can help reduce vulnerability by encouraging adequate, flexible coping mechanisms while discouraging those that are fragile or that limit economic and social mobility. Promising policies include creating self-regulating workfare programs and providing a supportive setting for institutions working to improve access to credit, crop and health insurance, and safe and convenient saving opportunities.

Many low-income countries, from Sub-Saharan Africa to Southeast Asia, have suffered major natural disasters and political upheavals in the 1990s. These events reminded observers of a reality hidden in official poverty statistics: that the condition of poverty is linked closely to vulnerability. Many poor households are exposed regularly to risks from illness, harsh weather, political instability, and economic mismanagement. Concern with vulnerability may be both intrinsic and tied to implications for the current generation as well as to the longer-term consequences for the health and education of children (Jacoby and Skoufias 1997; Hoddinott and Kinsey 1998; Rose 1999). Fear of risk can lead poor households to forgo potentially valuable new technologies and profitable production choices. Rosenzweig and Binswanger (1993), for example, use data from rural South India to show that an increase in risk (as measured by an increase of one standard deviation in the coefficient of variation of the

date of the onset of the monsoon) leads to a 35 percent reduction in farm profits [illegible] the poorest quarter of households but has no effect on the wealthiest farmers. V[illegible] nerable households may also spiral downward into poverty following adverse e[illegible] nomic or climatic shocks as productive assets are depleted to protect consumpti[illegible] levels. Addressing risk can thus be an important complement to redistributive [illegible] and antipoverty strategies focused on increasing economic growth and employ[illegible]

Yet even with holes in both public safety nets and private insurance markets, [illegible] households are not completely exposed to risk. Most have developed coping [illegible] gies to deal with the harshest blows. Most of these mechanisms are provided [illegible] by the market nor by the state but instead are private "informal insurance [illegible] ments. They include individual and community actions, such as drawing d[illegible] ings, selling of physical assets, reciprocal exchanges of gifts and loans, [illegible] crops, and expanding income-generating activities. (For recent surveys, se[illegible] man and Paxson 1994; Besley 1995; Morduch 1995; and Haddad and Zell[illegible] Some, like ritualized gift giving, have roots going back generations or even [illegible] while others are newer responses to difficult situations (Mauss 196[illegible])

Recent studies warn that some public policies may do little more than [illegible] these informal mechanisms, but most evidence shows that crowding out is n[illegible] to be a substantial problem. Most informal insurance mechanisms are typic[illegible] and often provide only inadequate protection to poor households. Studies from [illegible] gions as diverse as rural India, China, and Sub-Saharan Africa suggest th[illegible] informal insurance arrangements, households are exposed to considerable risk [illegible] adverse shocks—even idiosyncratic shocks that do not simultaneously affec[illegible] neighbors. Moreover, private informal mechanisms that are effective [illegible] vulnerability can retard economic growth and social mobility. Thus, even [illegible] informal insurance is well developed, public actions that displace inform[illegible] nisms can yield net benefits.

The emerging evidence suggests that policymakers need to be conc[illegible] more than providing disaster relief in the wake of large, aggregate sh[illegible] floods, earthquakes, droughts, and other natural disasters. It is equally [illegible] consider the needs of households that are facing losses due to adverse [illegible] nomic, or other crises such as illness, poor (local) harvests, and tempor[illegible] ployment. Policy options include creating a supportive environment for [illegible] that offer safe and reliable means for poor households to borrow and [illegible] to save. Recent experience shows that it is also possible to offer limited in[illegible] and protection against other basic exigencies in a simple, low-cost man[illegible] more speculatively, it may be possible to improve on existing insurance [illegible] ments for poor households by drawing lessons from the emerging micro[illegible] movement and by relying on nongovernmental organizations and profit [illegible] commercial enterprises to take key roles. Public workfare programs that offer [illegible] rary employment at wages that are too low to attract those who already ha[illegible]

income shocks but that the consumption levels of landless households and smallhol[illegible] decline sharply as income falls (Morduch 1993).

Gertler and Gruber (1997) take a different cut on tests of risk sharing. In [illegible] ing the ability to cope with the costs of illness in Indonesia, they find that [illegible] holds insure adequately against about 70 percent of common health shocks but [illegible] protect their consumption levels against only about 30 percent of the illness[illegible] seriously impair long-term performance. Their findings mirror Cochrane's [illegible] evidence using data from the United States and Lund and Fafchamps's (199[illegible] from the rural Philippines. Lund and Fafchamps, in particular, find that [illegible] insurance arrangements are effective only in the case of young adults who ar[illegible] ill; older adults who fall ill are far less likely to be helped. Lund and Fafchamps [illegible] show that informal insurance also helps with funerals but not with crop [illegible] mild illnesses, or unemployment (other than that of the head of the family [illegible] spouse). The household's social network also matters: households with mor[illegible] (especially richer friends) have a greater ability to use informal insurance [illegible] holds that are not so well connected fare much worse.

These results, based on disaggregations by class and type of shock, [illegible] nesses in informal insurance that would not otherwise be evident. And they [illegible] that there is ample scope for potentially beneficial interventions that go be[illegible] saster relief.

Informal Insurance Mechanisms

Reciprocal gift giving is a common way to solidify social and economic [illegible] —and one potentially important form of informal insurance. In North[illegible] example, gifts are given to mark births, deaths, and weddings, as well as [illegible] elderly, the ill, and women who have just given birth (Kipnis 199[illegible] [illegible] have tended to downplay gift giving as a product of a rational calculus [illegible] with informal insurance systems, instead highlighting its role in securing [illegible] tus and signaling commitment to the community (Malinowski 192[illegible] the sense of obligation engendered by gift exchange and the great poten[illegible] exchange to address risk, it is only natural that economists have take[illegible] insurance relationship seriously. Economists have thus tended to view gifts [illegible] do other transfers such as public aid (Cox and Jimenez 1991).

Private transfers of cash, food, and clothing are large and frequent in [illegible] tries. For example, 40 percent of black South Africans reported either [illegible] giving cash transfers (Cox and Jimenez 1997). Cash transfers reach large [illegible] urban residents in Colombia (Cox and Jimenez 1998), Thailand (Paulson [illegible] and the Philippines, where 82 percent of urban and 89 percent of rural house[illegible] report receiving transfers (Cox and Jimenez 1995).

Remittances from migrants—whether migrants to another country or to the city
om the country—can also be substantial (see Lucas and Stark 1985 on Botswana
nd Paulson 1995 on Thailand). Roughly two-thirds of all transfer inflows in Paki-
in between 1985 and 1988 came from migrants who sent money home to their
milies (Foster and Rosenzweig 1999). In the Philippines, 26 percent of urban and
percent of rural households received remittances from migrant parents or chil-
ren (Cox and Jimenez 1995).

But elsewhere, especially where migration is limited, reported transfers are of mi-
or consequence. In a data set from India fewer than 400 of 4,000 households re-
orted receiving net transfers in 1968–71 (Foster and Rosenzweig 1999). The lack
f transfers is especially notable in contexts in which they are expected to be most
aluable. In the smaller ICRISAT survey of poor villages in rural South India, Rosenzweig
1988) finds that transfers respond to risk but that they cover less than 10 percent of
he typical shortfalls in income. Studies of exchange in Sub-Saharan Africa by Reardon,
Dalton, and Delgado (1988) and Czukas, Fafchamps, and Udry (1998) reinforce
his picture. Data from rural China are only slightly more optimistic. In a four-year
tudy of 16 villages in the north of the country, Morduch and Sicular (1999) find
hat no more than a quarter of households report receipts of transfers from their
eighbors. 10 percent report receiving gifts from outside their village. The transfers
ere modest, moreover, averaging about 10–20 percent of average household
ncome.

Lim and Townsend (1998), who have examined ICRISAT files, conclude that house-
holds bridge the gap between income and desired consumption levels through sav-
ing, borrowing, and the use of buffer stocks of grain. The importance of these mecha-
nisms is evident in most low-income economies (Deaton 1992; Alderman 1996).
Households might also take actions to smooth income after an economic shock by
working longer hours, for example, or taking an extra job (Kochar 1999). And they
may take precautions beforehand to reduce the probability or extent of loss (Morduch
1995). These mechanisms can be relatively effective in the right circumstances, lead-
ing to concerns that publicly provided resources would simply replace—or crowd
out—these informal activities.

Reexamining the Costs of "Crowding Out"

To the extent that informal mechanisms are limited, the concern about crowding
out should be small. Still, some evidence is disturbing. Cox and Jimenez (1995) use
household level data from the urban Philippines to give a particularly striking ex-
ample. They estimate that the receipt of net transfers from other households is par-
ticularly sensitive to whether the recipient is unemployed. The magnitude of the
estimates is such that they conclude that if the government were to institute a simple

unemployment insurance scheme, net private transfers to the unemployed w[illegible] fall by 92 pesos for every 100 pesos offered by the public program. In the end, [illegible] average unemployed worker would be better off by only 8 pesos.

Exercises based on extrapolations like this yield provocative—but not definit[illegible] results. First, the research is not based on the effects of an actual program. Sec[illegible] researchers typically have data from just one side of any given transfer: who mak[illegible] or who gets it—but not both. Without complete information, it is difficult to t[illegible] out the exact reasons for the transfers. A transfer that looks like informal insur[illegible] against a bout of illness, for example, might instead reflect a correlation arising [illegible] other reasons, such as a gift from a child to provide a parent with old age support [illegible] the evidence from the Philippines, Cox and Jimenez (1995) concluded that the [illegible] cal retirement income variables reflected the desire of givers to help retirees. Is th[illegible] fact the case? Or do the variables instead signal a type of household (one with litt[illegible] no retirement income) that is more likely to include migrants—and thus to rec[illegible] remittances—rather than one without migrants? In the latter case, the receipt of [illegible] transfers could look like informal social insurance but instead could simply ref[illegible] steps taken to maximize household income. Disentangling the explanations requ[illegible] richer data.

Some of the sharpest evidence on crowding out is from South Africa. When [illegible] apartheid system was falling apart, the government extended basic pension bene[illegible] to black South Africans on terms similar to those that had been available to wh[illegible] who had no private pensions. The program, which was fully implemented by [illegible] 1993, provided a state pension equal to about $3 a day to all women over age [illegible] to all men over age 65, subject to a means test (Case and Deaton 1998). The m[illegible] test excluded nearly all whites and only the richest blacks. In the past, blacks had [illegible] to rely mainly on their own means to cope with aging and with economic d[illegible] turns, and for the most part the new benefits were not expected. How were [illegible] tional private mechanisms affected by the postapartheid public pension reform? [illegible] (1998) estimates that for those households receiving private transfers, every pub[illegible] provided rand led to a reduction of 0.2–0.4 rand in private transfers to the [illegible] Migration by children was also reduced slightly.

A similar degree of crowding out is predicted on the basis of studies of [illegible] urban residents in Peru and the Philippines: a displacement of [illegible] percent [illegible] percent, respectively, for each unit transferred as a retirement benefit (Cox and Jim[illegible] 1995, 1998). Without the chance to evaluate the introduction of an [illegible] program (as in Jensen 1998), the predictions derive from estimates of the [illegible] of net transfer receipts to existing social security or private pension arrangem[illegible] Cox and Jimenez then use coefficients from this exercise to predict the consequen[illegible] of the introduction of a broad state pension system.

Even with displacement rates as high as 20–40 percent, are the social losses [illegible] portionately high? Not necessarily, because leakage does not imply pure wastage[illegible]

The degree to which displacement undercuts policy objectives depends on the specific objectives of the program in question: Is it poverty reduction? Old-age support? Enhancing economic efficiency? Reducing vulnerability?

In the case of South Africa, older citizens received less in total than supposed, but others gained—and those gains were socially valuable. In richer contexts, private transfers tend to be from the old to the young. In a sample of white South Africans, for example, the net recipients were five-and-a-half years younger than the net givers (Cox and Jimenez 1997). But in poorer contexts, the reverse is most often true, largely because parents invest in children with the expectation that the children will support them in their old age. In line with that hypothesis, net recipients among black South Africans are on average eight years older than net givers. The displaced transfers thus tended to return to young households, many of which are as poor as older households, yielding little leakage as far as poverty reduction is concerned. In addition, keeping the funds in the hands of younger households is more likely to encourage investment in human capital accumulation and other productive activities. Second, public transfer systems may be more efficiently delivered than private transfers, yielding a net gain to society through displacement. For example, public transfer schemes may be able to pool resources more efficiently than local private arrangements (Cox and Jimenez 1997). Third, some displacement, even if it constitutes an unwanted leakage, may be a required cost of strengthening and widening the safety net to include particularly vulnerable households. In South Africa, for example, just under half of pension recipients do not receive private transfers at all (Jensen 1998). Putting the other arguments aside, tolerating some crowding out of the transfers received by half of the elderly black population can be seen as a cost of extending the safety net to the other half.

Tensions in Informal Insurance

W[illegible] evidence of risk sharing so weak in places where it is expected to be strong? [illegible] example, consider the highly risk-prone semiarid tropics of South India, where [illegible] of every ten years on average bring drought. Despite the importance of [illegible] weather, most of the variation in measured household incomes over time is [illegible] idiosyncratic to particular households. Morduch (1991) shows that 75–96 percent [illegible] variance of the logarithm of household income remains after removing variation due to changes in average village income and average household income over the study period (1976–82). Some of this idiosyncratic, residual variation is surely measurement error, but even if half of it is error, substantial idiosyncratic variation [illegible] remains. As a result, within-village gift exchanges, designed so that no net redistribution takes place over the period, could in principle reduce the variability of household after-gift income by as much as 90 percent in one of the villages

under study. In practice, reported transfers are not nearly big enough to do (Rosenzweig 1988).

Tensions in Systems of Reciprocal Transfers

To explain why systems of reciprocal transfers are weak, researchers point first problems in enforcing understandings. Household A will help household B and with the expectation that B will eventually reciprocate. But what will keep B from reneging? If A and B are related by blood or marriage, altruism may hold th together. But without altruism or enforceable contracts, self-interest is needed keep incentives in line. The repeated nature of the interaction over time allows self-interested reciprocity. If A can credibly commit to end all future insurance tionships with B in the event that B reneges, B may well see fit to fulfill obliga B's decision will depend on whether the gain from reneging today is smaller than flow of future benefits from continued participation.

As Coate and Ravallion (1993) suggest, however, the degree of effective insur that is provided will adjust so as not to tip the balance toward reneging. In pra tensions are heightened when both parties are down on their luck (during a dr say) or when a partner's luck is particularly bad. When an individual is pushed to the subsistence constraint, holding onto whatever one has may be especially te ing, despite the agreement to share with others. As a result, reciprocal exchange to fall apart (or to offer less of a return) when insurance is most needed. In gener works best when participants have a cushion against poverty. Consistent w evidence from China, the Philippines, and South India, theory suggests that syst of reciprocal transfers will be more effective for slightly richer households and in less dire contexts (Coate and Ravallion 1993; Ligon, Thomas, and Worr Kletzer and Wright 1998).

Moral hazard is also likely to limit group-based informal insurance, just as dermines standard insurance markets. When insurers cannot adequately obse enforce that insurees are taking all due precautions, incentives can be enha providing only partial insurance coverage. This is one reason that informal ins is most prevalent among relatives or neighbors in similar professions, tha with good flows of information.

Transfer-based systems can also run into trouble when households have o nities to accumulate savings because they then have a degree of insurance obligation to neighbors and kin. Similarly, when incomes of participants different rates, richer households tend to opt out rather than face the p systematically redistributing to others. Richer participants find themselves relatively more than they get back on average, and at a point they will leave form a new group with other richer households or to fend for themselves individu ally. Evidence of these sorts of bifurcations is given by Platteau and Abraham

relatively rich (who can cope better without group-based insurance) and the [illegible] tively poor (who never had much group-based insurance to start with) will migr[illegible] People in the large middle segment of the population will stay put, even thoug[illegible] may be economically beneficial for them to break their ties with the village and [illegible] the modern sector. The presence of informal insurance in villages can then be [illegible] on economic development. Drawing on data from Indian villages, Das Gupta [illegible] provides evidence along these lines. In parts of Sub-Saharan Africa the greatest pr[illegible] lems tend to be caused by excessive rural-urban migration rather than by insuffic[illegible] mobility, but even here the Banerjee-Newman model can still provide useful [illegible] sights. The basic ideas can be applied to explain inefficient mobility between [illegible] nomic sectors, for example, rather than just inefficient geographic mobility

The final set of tensions centers on the role of the family. The family has b[illegible] hovering in the background in this discussion because people generally turn to t[illegible] relatives first—and often again as a last resort—in times of need. On the one h[illegible] the institution of the family, stretching over generations and bearing well un[illegible] stood protocols, greatly facilitates informal insurance. Most important, inform[illegible] and enforcement problems are mitigated. On the other hand, the famil[illegible] have a much more limited pool of resources on which to draw relative to the broa[illegible] community.

The most important tensions arise when the demographic structure of housi[illegible] is shaped to meet the purposes of informal insurance. For example, the old age [illegible] rity theory suggests that children are produced partly to provide informal social se[illegible] rity. In situations with overcrowding and in cases in which parents do not tak[illegible] account the negative externalities imposed by their children (through conges[illegible] and environmental degradation, for instance), social welfare may be enhanced [illegible] shifting to alternative social security mechanisms (Dasgupta 1993; Arma[illegible] Morduch 1999). For example, establishing secure, convenient savings program[illegible] allow households to reduce the number of children they have without underm[illegible] their ability to cope with less income in old age and can provide a second [illegible] of benefits to the community through reductions in negative population [illegible] externalities.

A number of other insurance-demographic links lead to similar tensions [illegible] ing social pressure to migrate and to select marriage partners in order to prov[illegible] family with insurance (Rosenzweig and Stark 1989; Paulson 1995), family [illegible] ing" (that is, the turnover of responsibility) as a response to the death of [illegible] household; and the practice of taking in a foster child. Child fostering is comm[illegible] Côte d'Ivoire, Ghana, and Sierra Leone, although explanations differ. Bledsoe [illegible] Isiugo-Abanihe (1989) discuss insurance-related motives for fostering. Ainsw[illegible] (1996), however, in a survey from Côte d'Ivoire, finds that the need for labo[illegible] more important factor.

for girls than for boys. Reducing households' vulnerability and instituting more [illegible]ible insurance instruments may thus have broader social implications.

Policy Implications

A first set of policy priorities includes actions to reduce risk itself. For exam[illegible] improving governance can sharply reduce the vulnerability of households to [illegible] turns resulting from economic mismanagement. Increasing macroeconomic [illegible]ity, reining in inflation, securing property rights, improving transport and [illegible]nications, and creating a stable political environment can go a long way [illegible] reducing the frequency and size of downturns and creating a supportive [illegible]ment to facilitate private risk-reducing activities. Similarly, risk can be reduc[illegible] public health campaigns for immunization and sanitation, civil works projects [illegible] retaining walls, irrigation), and, in some cases, price stabilization. Higher [illegible] and stable employment opportunities further enhance the ability to cope [illegible] But these are all policy areas that are on the table for other reasons [illegible] judged by other criteria.

In richer countries, households typically prepare for income declines by [illegible]ing savings accounts, lines of credit, pensions, insurance, and annuities. [illegible] these actions run into limits, governments typically provide means-tested [illegible] alleviation programs, unemployment benefits, health insurance, and social [illegible]rity (Subbarao and others 1997, ch. 3). But neither the administrative cap[illegible] the funding exists in most low-income countries to build similar public [illegible] Public action can, however, help to address smaller, local hardships by [illegible] regulatory and institutional frameworks that expand households' access [illegible]ance, credit, employment opportunities, and convenient ways to save. [illegible] the government's role conserves scarce administrative resources and [illegible]tial conflicts of interest between short-term political exigencies and [illegible] for longer-term institutional sustainability. These policies provide ways [illegible] informal coping mechanisms and broaden their accessibility rather than [illegible] private actions.

Promoting Savings

It had long been thought that most poor households had little desire [illegible] banks, but the experience of Indonesia's Bank Rakyat Indonesia (BRI) and [illegible] microfinance programs are turning that view around. After BRI established [illegible] convenient savings vehicle, consumers responded enthusiastically. BRI now has [illegible] than 16 million low-income depositors (compared with 2 million borrowers), [illegible] aiding the bank's profitability. Although there is no systematic evidence on [illegible]

ne levels of depositors, bank staff note that they tend to be poorer on average than
rowers and from diverse socioeconomic backgrounds. Partly as a result, savings
bilization efforts are now being renewed in microfinance programs in Africa, Asia,
Latin America. Public policy can aid by ensuring an appropriate regulatory envi-
nment and helping to keep inflation in check.

One promising program has shown the surprising demand for savings deposits
ong poor households in the slums of Dhaka, Bangladesh. SafeSave, a nongovern-
ntal organization, patterned its savings program on that of the local rotating sav-
g and credit associations, which operate by taking in small sums from participants
utherford 1999). The response to SafeSave has been much greater than expected,
d depositors have been able to slowly build up usefully large sums of money. As a
nsequence, the program is being replicated by other nongovernmental organiza-
ns in South Asia.

Without easy saving opportunities, households are tempted to squander surpluses
are susceptible to calls for short term help from family members or neighbors—
en at the expense of long term progress (Platteau forthcoming). In this way, sav-
gs instruments may well be much more important than the provision of credit in
sing incomes and reducing risk — and easier to accomplish.

Such financial deposits can be particularly effective in helping households weather
difficult scenarios that undermine gift exchanges. Consider the various forms of
ks that households encounter. Events that occur infrequently—old age, death in
e family, and chronic disability — can hit households hard and may require a con-
ing flow of transfers. Given that such transfers are not guaranteed, savings de-
sits can be critical in ensuring that people have enough income to satisfy basic
eds. Note, however, that if interest rates fall below inflation rates, the purchasing
ower of these deposits can erode quickly. Savings also allow households to avoid
rowing from moneylenders at interest rates as high as 5–10 percent a month
en emergency funds are needed and can be especially valuable in a crisis (Von
ke 1991). A regional drought, for example, will lead to a decline in the price
assets as affected individuals simultaneously try to sell their holdings. Financial
s, however, will generally hold greater value (and could increase in value as
all).

lic policy that leads to better integrated savings programs can help to contain
allowing the financial system to handle shocks more easily. The theoretical rela-
ship of deposit mobilization, efficiency enhancement, and the generation of eco-
growth is described by Bencivenga and Smith (1991). The keys to a successful
gs program are providing long-term security and convenience, finding a way to
dge against inflation, minimizing costs, and exploiting opportunities to relend
posits safely but profitably.[5] Existing banks and nongovernmental organizations
may not be up to all of these tasks, and designing effective (but not overly intrusive)
prudential regulations is a critical first step.

[illegible]

Microcredit

Microcredit programs have succeeded by creating hybrid institutions that cha[illegible] formal-sector funds to poor households. The programs are not perfect. In Africa, [illegible] example, the challenge is to create mechanisms that work well in semiarid and rural regions where households tend to have less diversified income bases and w[illegible] low population densities mean higher transactions costs for financial institu[illegible] More generally, most poverty-focused programs face high costs that undermine [illegible] tempts at profitability. A recent survey shows that such programs are able to cover average of only 70 percent of their total costs (*MicroBanking Bulletin* 1998,

The benefits, though, may be considerable. Using a recent survey of 1,800 ho[illegible] holds in Bangladesh, I find that access to microcredit programs yields no appre[illegible] increase in average consumption levels in the short term (Morduch 1998). For [illegible] with access, however, the volatility of consumption over the three main cropp[illegible] seasons is roughly half that of control groups (after controlling for unobse[illegible] variables at the village level). This reduction in consumption variability turns [illegible] be mainly a product of reduced income variability across seasons, which is m[illegible] possible by the employment diversification that credit affords. Helping rural [illegible] holds to reduce risk further by diversifying into nonfarm labor is an overlooke[illegible] important, return to microcredit, and more research needs to be done along [illegible] lines to inform discussions of the costs and benefits of supporting credit-base[illegible] proaches (Khandker 1998; Morduch forthcoming).

Insurance

Crop insurance programs have been a disaster nearly everywhere [illegible] not unlik[illegible] geted credit programs in the 1970s (Yaron, Benjamin, and Piprek 199[illegible] [illegible] information and high transactions costs have proven to be destabilizing [illegible] are no easy solutions. Although reform currently looks unpromising, in princ[illegible] problems of insurance markets are not much more intractable than those [illegible] markets—and microfinance programs have shown effective ways around [illegible] the largest hurdles there. Some microfinance programs have introduced [illegible] successfully on a limited scale, offering term life insurance at very low [illegible] benefits large enough to clear debts and provide for a burial but not much [illegible] addition, the Grameen Bank of Bangladesh, for example, appears to have [illegible] cess with its "emergency fund" for borrowers. The fund aids with loan [illegible] and provides general help in the event of illness and other emergencies. Inform[illegible] and transactions costs are reduced by coupling these mechanisms with credit [illegible] sion. Experimentation will be necessary to determine whether these types of [illegible] ance mechanisms can be provided separately from other microfinance serv[illegible]

As Morduch and Sicular (1999) suggest, there may be ways to insure poor house-ids by drawing broader lessons from microfinance. In a study of northern China, describe an insurance company that has found some success in selling crop insur-e to groups (a whole village, say), rather than to individuals. At the moment, this up insurance is used just to lower transactions costs for insurers and is a poor logue to the group-lending practices used by microfinance institutions. But if ure premiums were tied to the history of losses, a group-based contract could vide incentives for peer monitoring along the lines that microcredit programs e found successful in addressing moral hazard. This is an area open to specula-n, and many roadblocks remain — for example, how should a program discourage llusion by the entire group?

One important lesson from microfinance is that programs operated directly by vernments tend to have inherent difficulties in generating compliance by partici-nts; borrowers are far more likely to default on loans from government sources, d governments are more likely to tolerate defaults in the name of political expedi-cy. This has proved disastrous for the long-term sustainability of public credit ograms. There is a parallel in the case of insurance: insurees appear less likely to ke due precautions when governments are the insurers. Facilitating insurance pro-sion by nonprofits, nongovernmental organizations, and for-profit companies may thus be an important step forward.

mployment Guarantee Schemes

ect public interventions can also help to reduce vulnerability, especially for the rest households. Among the most promising are rural public works programs h as India's Employment Guarantee Scheme in Maharashtra State, mentioned e, and the Food for Work program in Bangladesh, both of which are described Ravallion (1991). These programs provide wage employment in return for work constructing and maintaining public infrastructure. Ravallion (1991) reports that e scheme provided about 100 million person years of employment between and 1989. On average, 500,000 people participated per month (of a total of 20 agricultural workers). Walker, Singh, and Asokan (1986) find that the coefficient ariation of income among landless laborers in two villages in Maharashtra with ess to the employment scheme was half that of a similar village without access.

The work requirement and low wage rates provide a way to target the aid to needy households, allowing the programs to avoid instituting costly means . Participants take advantage of the program only when needed, often in lean ons before harvests. During peak seasons, alternative employment opportuni-s are generally more attractive. Thus the programs avoid long-term dependency d ever growing lists of participants. It also means that the programs are set up to

help households cope with temporary hardships, not mainly as an answer to chr[illegible] poverty.

The success of such programs will depend on government budgets, eligibility [illegible]teria, and wage rates. According to Ravallion (1991:171–72), there should be [illegible] few restrictions on eligibility as feasible, and wage schedules and the rights of [illegible] pants should be well defined, well known, and nondiscriminatory. Ideally, [illegible] want work at the going wage rate should be able to get it." The principle of inclusi[illegible]ness is a key to reducing vulnerability because households (or at least those [illegible] available workers) are reassured by knowing that they will have a place to turn [illegible] they fall on hard times.

Conclusion

Poor households throughout the world face twin disadvantages. The first is diffic[illegible] in generating income. The second is vulnerability to economic, political, and [illegible] cal downturns. Inflation, recession, drought, flood, illness, and civil war hit [illegible] those households that are least well equipped to handle the shocks. Harder still [illegible] two disadvantages reinforce each other. Poverty is a source of vulnerability [illegible] peated exposure to downturns reinforces poverty.

The circular nature of poverty and vulnerability does not, however, preclude [illegible] fective responses. The evidence to date suggests several broad directions to pursue [illegible] addition to helping households cope with large natural disasters, governments [illegible] to encourage flexible private interventions. Concern has arisen about whether [illegible] action will crowd out private informal insurance mechanisms. To the contrary, [illegible] designed public action can strengthen and broaden the capacity of households [illegible] independently through informal mechanisms.

Making saving safer and more convenient, helping to expand credit access [illegible] fostering basic insurance programs are all promising ways to help households [illegible] themselves in the face of adversity. The possibility of crowding out existing [illegible] arrangements should not be ignored, but in most low-income countries [illegible] likely to substantially undermine steps to help poor households. First, inform[illegible] surance is often very limited, and second, the crowding out of some private [illegible] can have valuable social benefits.

Notes

Jonathan Morduch is a MacArthur Foundation Research Fellow at Princeton University. [illegible] is grateful for comments from Harold Alderman, Angus Deaton, and the journal's referees. [illegible] also benefited from discussions with Marcel Fafchamps, Frima Haque, John Hoddinott [illegible]

Malinowski, Bronislaw. 1922. *Argonauts of the Western Pacific*. London: Routledge.

Mauss, Marcel. 1967. *The Gift*. New York: W. W. Norton.

MicroBanking Bulletin. 1998. Boulder, Colo.: Economics Institute. July.

Morduch, Jonathan. 1991. "Consumption Smoothing across Space." Harvard University, Department of Economics, Cambridge, Mass. Processed.

———. 1993. "Risk, Production, and Savings: Theory and Evidence from Indian Hou[illegible] Harvard University, Department of Economics, Cambridge, Mass. Processed.

———. 1994. "Poverty and Vulnerability." *American Economic Review* 84(May):221–25.

———. 1995. "Income Smoothing and Consumption Smoothing." *Journal of Economic P[illegible]* 9(Summer):103–14.

———. 1998. "Does Microfinance Really Help the Poor? Evidence from Flagship P[illegible] Bangladesh." Stanford University, Hoover Institution. Stanford, Calif. Processed.

———. Forthcoming. "The Microfinance Promise." *Journal of Economic Literature*

Morduch, Jonathan, and Terry Sicular. 1999. "Risk and Insurance in Transition: Perspe[illegible] Zouping County, China." In Masahiko Aoki and Yujiro Hayami, eds., *Market and* [illegible] *Economic Development*. New York: Oxford University Press.

Newbery, David, and Joseph Stiglitz. 1981. *The Theory of Commodity Price Stabiliza[illegible]* Oxford University Press.

Paulson, Anna. 1995. "Insurance Motives for Migration: Evidence from Thailand." [illegible] versity, Princeton, N.J. Processed.

Platteau, Jean-Philippe. 1991. "Traditional Systems of Social Security and Hunger Insura[illegible] Ahmad, J. Drèze, J. Hills, and A. Sen, eds., *Social Security in Developing Countries*. Oxford [illegible] Press.

———. Forthcoming. "Traditional Sharing Norms as an Obstacle to Economic Growth [illegible] Society." In *Institutions, Social Norms and Economic Development*. Chur, Switzerland: [illegible] Academic Publishers.

Platteau, Jean-Philippe, and Anita Abraham. 1987. "An Inquiry into Quasi-Credit C[illegible] Role of Reciprocal Credit and Interlinked Deals in Small-Scale Fishing Communities." *Development Studies* 23(4):461–90.

Ravallion, Martin. 1991. "Reaching the Rural Poor through Public Employment: Arg[illegible] dence, and Lessons from South Asia." *World Bank Research Observer* 6(2):15[illegible]

Ravallion, Martin, and Shubham Chaudhuri. 1997. "Risk and Insurance in Village [illegible] ment." *Econometrica* 65(1):171–84.

Reardon, Thomas, Peter Matlon, and Christopher Delgado. 1988. "Coping with H[illegible] Food Insecurity in Drought-Affected Areas of Burkina Faso." *World Development* [illegible]

Rose, Elaina. 1999. "Consumption Smoothing and Excess Female Mortality in Rur[illegible] *of Economics and Statistics* 81(February):41–49.

Rosenzweig, Mark. 1988. "Risk, Implicit Contracts, and the Family in Rural Areas of [illegible] Countries." *Economic Journal* 98(December):1148–70.

Rosenzweig, Mark, and Hans Binswanger. 1993. "Wealth, Weather Risk, and the Com[illegible] Profitability of Agricultural Investments." *Economic Journal* 103(January):56–78.

Rosenzweig, Mark, and Oded Stark. 1989. "Consumption Smoothing, Migration [illegible] Evidence from India." *Journal of Political Economy* 97(4):905–26.

Rosenzweig, Mark, and Kenneth Wolpin. 1993. "Credit Market Constraints, Consumptio[illegible] ing, and the Accumulation of Durable Production Assets in Low-Income Countries: [illegible] in Bullocks in India." *Journal of Political Economy* 101(April):223–44.

Using Household Surveys to Build Analytic Capacity

Lorraine Blank • Margaret Grosh

[illegible]ticle reviews the results of efforts in five countries to build a national capacity to [illegible] social policy. these efforts were undertaken in conjunction with Living Standards [illegible]ement Survey (LSMS) projects. Case studies for Bolivia, Jamaica, the Kyrgyz Re[illegible], South Africa, and Vietnam show that building such capacity requires explicit [illegible] significant time and money, open access to data, and strong support from [illegible]. Lessons are drawn about four aspects of building capacity —training, tech[illegible] assistance, research, and recruitment. The lessons learned from these LSMS projects [illegible] more general sense to other kinds of capacity building projects.

[illegible]proving the effectiveness, efficiency, and responsiveness of government is a [illegible] for developing countries as they prepare for the 21st century. The ability to [illegible] the impact of economic and social reforms, evaluate the outcomes of public [illegible] and projects, and undertake policy-relevant research is an important com[illegible] of good governance, and together these institutional capabilities constitute [illegible] referred to as the capacity for policy analysis.

[illegible] analysis uses multiple methods of inquiry to examine the quantifiable out[illegible] public policies to determine if the goals and objectives of a particular policy [illegible] are being achieved (Waterman and Wood 1993; Dunn 1994). Such [illegible] is an important input into the process of formulating and evaluating gov[illegible] policies. The field is grounded in economics but draws upon other disci[illegible] well. The techniques used include quantitative modeling, statistical analy[illegible] metrics, qualitative research, and political and institutional analysis. This [illegible] summarizes the results of efforts in five countries to build a national capacity [illegible] social policies covering welfare, education, health, fertility, nutrition, la[illegible] and employment.

[illegible] national capacity to undertake such operations is weak in most develop[illegible] countries, the current practice is to rely largely on foreign researchers (contracted

[illegible]

by the government through externally funded projects or directly by internatio[illegible]
development agencies). We believe that a strong case can be made for generat[illegible]
good policy analysis locally, whether commissioned by the government or conduc[illegible]
by in-country analysts. Commissioned analysis is done at the time and on the [illegible]
that the government requires, and local researchers who know the environment [illegible]
more likely to take into account domestic institutional issues and may be more kn[illegible]
edgeable about the full range of data that could be brought to bear on an issue. [illegible]
integration of the results into the policymaking process requires a sense of owners[illegible]
of the data and analysis that comes only with being part of the process from beg[illegible]
ning to end. Local researchers may be better placed to disseminate the results [illegible]
provide follow-up advice on their studies. Involving national policymakers and [illegible]
lysts in the process of setting the agenda for policy analysis and organizing the [illegible]
collection to support it ensures that lo[illegible] needs are met. This means that build[illegible]
local analytic capacity should produc[illegible] effective social policies.

Several factors have contributed t[illegible] limited analytic capacity in many devel[illegible]
ing countries. Problems with the qu[illegible] of the data are extensive. Where adequ[illegible]
data have been collected, opportunities for their use are often missed because [illegible]
searchers are given limited— or no— access. And countries often have very few [illegible]
searchers, especially those connected with established universities and trained [illegible]
vanced statistical techniques and research methodologies. Although access [illegible]
computers and software may be adequate, there is likely to be little in-country ex[illegible]
rience with statistical packages. Graduate training in policy analysis is either no[illegible]
istent or lacking in rigor. Finally, but most significantly, there is no culture of [illegible]
social data to address policy questions. Policymakers have little experience [illegible]
policy questions in ways that are amenable to analysis; they also are frequently [illegible]
willing to provide open access to information. As a result government offic[illegible]
dom demand policy analysis (Peterson 1991; Simonpietri and Ngong 199[illegible]
1996; Thorbecke 1996). These factors are related in a vicious circle. If data [illegible]
collected or not released to analysts, aspiring students will not hone their skills [illegible]
quantitative analysis but will move into other areas of their disciplines. If policym[illegible]
do not have access to good policy studies, they do not learn to integrate [illegible]
the policy process, and they are not interested in funding data collection [illegible]

To address these limitations, international development agencies, includin[illegible]
United Nations, the U.S. Agency for International Development (USAID) [illegible]
World Bank, have provided financial support for household and regional [illegible]
and other efforts to collect quality data by providing technical assistance [illegible]
and computer hardware and software to policy analysis units (both independent [illegible]
in government), research centers, and training programs. Other efforts to [illegible]
capacity have involved using local researchers in externally funded research [illegible]
and training programs (Paul, Steedman, and Sutton 1989; Grindle and Hilderbr[illegible]
1995; Myers 1997), establishing regional training programs and peer review [illegible]

was not designed to build analytic capacity at all but is typical of projects wh[illegible] single analytic imperative drives a survey project. The first Vietnam project was [illegible] larly focused on data collection but thoughtfully used small amounts of mone[illegible] capacity building. In South Africa another single-survey effort did involve a [illegible] number of potential future analysts in the survey design and dissemination pla[illegible] contrast, Bolivia and Jamaica were multiyear, multimillion-dollar efforts design[illegible] build analytic capacity.

Kyrgyz Republic

The first Kyrgyz Multi-Purpose Survey (KMPS) was implemented in 1993 to [illegible] the government and other local institutions assess the poverty and employm[illegible] ation in order to design a program of targeted social assistance. The goal [illegible] collect national data as quickly as possible. No budget was specifically allotted [illegible] building analytic capacity.

At the time, the country's existing capacity was limited to implementing [illegible] dom household surveys that were designed and subsequently analyzed in Mos[illegible] Local researchers were not trained in quantitative analysis and had no experie[illegible] analyzing large data sets, either in government or in academic institutions. Com[illegible] ers were used, and programmers were available but were unfamiliar with stat[illegible] programs. Policymakers did not know how to use the survey to inform policy de[illegible] sions. Driven by the need to design a social safety net project as quickly as poss[illegible] the survey project did not include building local capacity for data collection [illegible] sis. The construction of a sampling frame would, however, provide the sam[illegible] basis for future collection of survey data.

The results of the analysis were presented at a seminar in September [illegible] tended by 50 senior officials from government ministries, trade unions, and [illegible] versities. Participants were invited to request specific tabulations related to [illegible] individual spheres of interest, and these were prepared and distributed the fo[illegible] ing day. The presentation of these customized tabulations generated conside[illegible] excitement among the attendees, who recognized the potential contribu[illegible] such analysis to policy decisions. This experience helped pave the way fo[illegible] ond, multiyear project with explicit goals for building analytic capacity. This [illegible] funded four rounds of the survey—two in 1996, one in 199[illegible], and one in [illegible] formal user committee representing the social sector, economic ministr[illegible] Goskomstat (the Kyrgyz statistical agency) reviewed the content of each quest[illegible] naire. The Research Triangle Institute, an independent nonprofit research [illegible] development organization headquartered in North Carolina, provided tech[illegible] assistance, including training in data collection and analysis. Experts worke[illegible] staff from Goskomstat and government ministries to produce poverty profile[illegible] statistical abstracts. Selected staff from these agencies participated in short [illegible]

nd overseas training programs in survey research, data processing and analysis, onstruction of consumption aggregates, poverty measures, anthropometric measurement, and sampling theory. The project cost $3.7 million, of which $545,400 as for training.

Was the program successful? Although local staff participated in the design of ach survey, their contribution was limited by their lack of experience with the uses f survey data. As anticipated, they were more active in the production of the 1996, 997, and 1998 poverty assessments and helped to develop pension models based on ie 1996 survey.

Two factors have limited the development of Kyrgyz analytic capacity: first, the istical methods that had been in place for 70 years were considered perfectly lequate by policymakers, and second, Kyrgyz analysts were suspicious of the household survey because it had been designed, processed, and analyzed in a foreign coun- . Despite the declaration of open access to the data, researchers were not familiar nough with the new instrument to forsake traditional data sources. Thus it was ard to get people to demand training to use that data. Moreover, policymakers have ot requested much information collected from the surveys. Although efforts to stimulate demand were made immediately after the first survey, they have had little lasting npact. Observation missions to other countries are scheduled, but these efforts are oming late in the project. Finally, it is generally agreed that in the rush to implement and analyze the surveys, technical assistance has been directed more at completing analytic tasks than at transferring analytic skills — a common phenomenon in any projects.

South Africa

The 1993 survey was undertaken to create a credible national database on poverty se in monitoring poverty, analyzing policy, and designing programs. A related ctive was to strengthen the capacity for these tasks among individuals who were to join a new majority rule government. The African National Congress and ongress of South African Trade Unions requested the project, which was conducted outside of government. The Labor and Development Research Unit at the versity of Cape Town coordinated the project. Originally envisaged as a two- e project, the first step focused on upgrading and filling the gaps in national ics information and the second phase on strengthening the capacity for continuous poverty monitoring. In the end, only the first part was implemented. The ect cost about $1.5 million, of which approximately $500,000 went to activities closely linked with building analytic capacity, including workshops, regional ies, and publication of the survey abstract.

The survey questionnaire was designed in a series of workshops attended by more 130 social scientists over a one-year period. Most of the fieldwork was carried out

[illegible]

in the latter part of 1993. The data were available for the new majority-rule go
ment in 1994.

Although there are 22 universities in South Africa, and many researchers been educated abroad, few individuals have been trained in quantitative ana As a result, there were few analysts, either in government or in academic ins tions, who were able to analyze household data sets. This was especially true black South Africans. Three workshops were held to train researchers in pro ing and analysis, and training in reading the data was provided to staff memb several government ministries in 1995 and on a demand (and fee-paying after that.

Access to data was completely open, and a very active data dissemination prog was implemented. Within a year after the survey was completed, data sets had distributed to 13 government agencies, 10 universities (7 of the 10 were tradit white), 9 independent research institutes, international aid agencies, and local and international individuals. By 1996 more than 36 research report had produced from the survey. The findings were used to inform government de regarding unemployment programs for women and pension programs for el erly; a white paper on water and sanitation, and a report on key indicators of p And the data were being used in university courses to strengthen interest and policy analysis.

As a result, the country's capacity to implement a national household sur proved in a "learning by doing" framework with technical assistance from Bank staff in questionnaire design, sampling techniques, fieldwork, data pro and preparation of the statistical abstract. The capacity to analyze the data information such as that contained in the statistical abstract was enhanced process that began with debating many versions of the questionnaire gions of the country. This experience provided an orientation for social who had not previously worked with data from an integrated household su it supported the Central Statistics Office's efforts to undertake household that more or less duplicate the objectives of the LSMS survey

Policy and research analysts in South Africa began talking about household sets and the analytic capacity necessary to use them. As a result capacity efforts have been included in a three-year Mellon Foundation project that stream in 1998. The goal of the project is to increase the number of social s policy analysts, and others with the skills needed to design, implement, and data. The project will fund graduate work in demography at the University Town, including long- and short-term training in quantitative social science meth ods as well as the West Cape Area Study, which will be implemented with assistance from the University of Michigan.' The area study will be used as a tool for graduate students as well as a mechanism to generate reliable data project costs will amount to $330,000.

The challenge is to respond to the demand for increased quantitative analysis by lping the government establish a more permanent system for monitoring poverty d to increase the capacity for policy research and policy analysis in relevant gov-nment agencies and in the less established and primarily black universities.

etnam

e first Vietnam Living Standards Survey was conducted in 1992–93 with funding om the United Nations Development Programme (UNDP). The goal was to gener-e a comprehensive data set to help guide the policy decisions being taken as part of e transition to a market-based economy. The project was jointly managed by the ate Planning Committee and the General Statistics Office. The costs of imple-enting the survey were $516,000, including $80,000 for training.

At the time, the country's analytic capacity was limited to basic statistics, with tle expertise in advanced analytic techniques. Where expertise existed, reliable raw ta and computer hardware and software were absent. Capacity-building strategies the first project focused on preparing and disseminating the published abstract; a udy mission to Thailand to observe survey procedures and to look at how the data re utilized; and short-term training (two to three weeks) in the use of statistical ftware, the structure and contents of the data set, and policy analysis. The project pported learning by doing through preparation of the survey abstract and working pers based on the data. The Ford Foundation funded a six-week workshop in licy analysis, statistics, computer training, preparation of research papers, and an itation to the data, followed by a two-week writing workshop on preparing re-h papers. The Ford Foundation provided participants in this workshop with a puter powerful enough to use the survey data, a copy of a statistical software age, and a copy of the survey data. Project funded training targeted policy ana-and technical staff from the central government, including the General Statis Office, the State Planning Committee, and social sector ministries; Ford Foun-n training targeted researchers from universities and research centers outside e government. The Ford Foundation provided $ 0,000 for training and equip-t and $50,000 to sponsor local analysis.

he first survey resulted in some positive changes in the policy analysis environ-t. To understand the extent of these changes, it is important to remember that the y broke new ground in Vietnam in several ways: it was the first household survey ut the unit record data on computer, it was the first survey that included training tions for ministry staff, and it was the first time that the statistical package had used in Vietnam. That the statistical abstract was produced, largely by Vietnam-nalysts, just six months after the data were collected was a significant achievement. he initial survey and surrounding activities created demand for further training data analysis. A second survey was fielded in 1997–98 as part of a project to

develop capacity to collect, analyze, and disseminate social statistics data. The pro
is funded by the UNDP and the Swedish International Development Agency
with total project costs of $1,782,400, of which $700,000 is earmarked for trai
The project is funding study tours, short-term in-country training in statistics
data analysis for participants from 22 agencies, and short-term overseas training
basic and advanced statistics, survey design and planning, sampling theory and p
erty measurement.

A wide range of international researchers and donor agencies have requested
data from the first survey, but the Vietnamese government itself has used the d
only minimally. That is partly because analytic capacity is still limited but prin
because there is as yet little demand for data-driven policy analysis. Participants
training programs are rarely called upon to use their newly acquired skills once
return to their jobs.

Training has been extensive in terms of the number of institutions reached
only a limited number of individuals within each organization is involved, and
training is fairly superficial. In the absence of long-range planning geared to dev
ing the skills required to implement and analyze household surveys, st
build these skills progressively have been lacking. Consequently, training h
offered to as many participants as possible instead of being provided in dep
more limited number of participants. The workshops give trainees the basic
cal skills necessary for descriptive analysis, but the skills required to formul
search questions and prepare research reports have yet to be acquired. The chal
for the future are to provide long-term training to participants who have be
lected on the basis of ability, rather than seniority, and to develop a strat
ensuring an ongoing supply of policy analysts.

Bolivia

The first LSMS-type survey was conducted in Bolivia in 1989. Five annual s
followed, at a total cost of $2.6 million, including $1.5 million for trainin
ing was directed primarily at implementing and processing the surveys, and
the changes that were adopted were applied in the government's subsequen
hold surveys. Despite the existence of the LSMS survey data and data fr
sources, there was little interest in social policy analysis. Bolivian universit
have a tradition of quantitative policy research. Similarly, there was ver
lytic capacity in the social sector ministries; for example, researchers who
analyze Statistical Institute data would ask for tables rather than the raw da
nal consultants conducted most of the studies that were done on household
data.

By the late 1980s, however, as social policy issues rose to prominence
national political agenda and the work programs of the international do

ncies, the gaps in in-country analytic capacity became obvious. As a result USAID lertook a $2.4 million project to build up the country's domestic capacity ough the establishment of a social policy analysis unit known as UDAPSO. The ernment contributed to operating costs with funding from other USAID-sported programs. SIDA paid for five staff positions through the Bolivian civil vice program. The group included a multidisciplinary team of up to 12 profes-nals and concentrated on education, health, poverty, income distribution, and ial sector expenditures.

The Harvard Institute for International Development (HIID) provided 32 months long term technical assistance and almost 40 months of service of short-term con-tants. These consultants advised on best practices, interpretation of results, and fu-e work and provided feedback on the quality and relevance of the work. In-house ining (in the form of two-week workshops on the use of large data sets and statistical tware, basic and intermediate statistics, econometrics, and other analytic tools) was o provided. Staff participated in workshops at Harvard in more general areas such as erty measurement, education planning, and health planning. On-the-job training urred through participation in research projects and preparation of research re-ts. The World Bank also provided 1 month of training in econometrics.

Between 1992 and 1995, these newly trained analysts produced 5 books, 2 re-rch monographs, and 36 working documents. LSMS data were essential to a great al of this research. The work was well received by other government agencies and ernational donors. Staff members participated in policy dialogues with senior gov-nment officials, and the dissemination of research allowed analysts to develop col-orative alliances with multilateral organizations.

The policy analysis unit's early success resulted from a combination of intangible tors including political support from the minister of planning, leadership, and rit de corps. When the minister and director of the unit changed, most of the fessional staff left, and recruiting difficulties plagued the organization. It was dis-nded in 1997. Despite this setback, the government did not abandon its capacity-lding commitment but instead created a division for social policy within the Min-y of Planning based on many of the former unit's functions and staff. This move y actually increase the impact of previous capacity-building efforts because the v division is closely linked to the Ministry of Finance, providing it with the tech-al analysis that underlies budget decisions. As a result, it has achieved legitimacy ong the sector ministries.

To ensure a continuing supply of trained policy analysts, USAID funded a project 1994 to establish a new master's degree program in public policy and manage-nt at the Catholic University of Bolivia. HIID has provided technical assistance for orks with faculty and staff in managing the program, teaching courses for graduate tudents and for public and private sector managers, and acquiring library and teach-ng materials.

[illegible]

Jamaica

The Survey of Living Conditions (SLC) was designed to collect data to monitor social impacts of public policies and the delivery of social services. The survey [illegible] first implemented in 1988 and has been carried out annually since then. By [illegible] SLC Jamaica's capacity for data collection was carried out in a smooth and [illegible] fashion and was providing an exceptionally rich database. In-country analysis [illegible] ever, was limited to production of the annual abstracts and some studies [illegible] quality) produced by local academics; few in-country researchers were [illegible] undertaking sophisticated analysis.

With the assistance of the World Bank and funding from the Netherlands [illegible] project was developed to increase the analytic capacity of the Planning Institute, sector ministries, and the University of the West Indies. The $3.4 million project [illegible] on stream in July 1993. About $2.8 million was devoted to building analytic [illegible] and the rest to conducting further rounds of the SLC. The project established [illegible] policy analysis unit in the Planning Institute; outlined a program of research; [illegible] provided technical assistance for graduate courses in statistical analysis, research methodology, and policy analysis for senior researchers in the sector ministries and [illegible] ulty at the university, as well as graduate school curriculum development. [illegible] set up a social indicators data bank; graduate fellowships; workshops in [illegible] search methods, techniques of policy analysis, and computer processing for [illegible] search staff in the sector ministries; and seminars to introduce policymakers to [illegible] of data in policymaking. It also funded computer equipment and technical [illegible] to the Ministry of Labor and Welfare (the only ministry not receiving assistance [illegible] other donors) and support for five rounds of the SLC.

The social policy analysis unit, known as the Policy Development Unit [illegible] tablished and operational for the whole life of the project and serves as [illegible] and a project implementation unit. The unit has produced policy studies [illegible] and has developed a methodology for tracking poverty over time. [illegible] and updated the poverty line studies that the government had [illegible] 1989, and they played an important role in preparing the government [illegible] strategy. The SLC abstract is now produced solely by in-country analysts. [illegible] ness and complexity of the abstract have increased since the project's inception.

When the Policy Development Unit was first established, staff salaries were [illegible] petitive with those in the public sector. It was difficult to find qualified [illegible] to fill the analyst posts because the pool of trained individuals was [illegible] Salaries have eroded since then, relative to government pay scales, and are [illegible] competitive. The turnover of professional staff has been high. Although [illegible] number of trained individuals is available (as a result of the project), recruitment [illegible] been difficult. This is a reflection not only of budget constraints but also [illegible] diminished political support for the unit.

The University of Toronto provided technical assistance to upgrade graduate education in social policy analysis at the University of the West Indies. Ten-week courses provided training in statistics, multivariate analysis, research techniques, cost-benefit and effectiveness analysis, and other techniques to graduate students, faculty, and junior researchers from the sector ministries and the Planning Institute. The series was run twice. Eighteen fellowships were awarded, and a much larger cadre of graduate students participated in the courses. The courses have been made part of the permanent curriculum, but their continued delivery is not guaranteed because none of the current faculty has the required skills to teach the quantitative courses. The university is making a serious recruitment effort, but its success is not assured.

A social indicators data bank, which houses a computer laboratory and a library of important social sector data sets, including SLC data, was established at the university, which now covers the recurrent costs of the data bank. The data bank is encouraging the integration of hands-on data analysis in a wider range of classes than heretofore possible and is thereby encouraging quantitative work in universities more broadly.

A series of in-service workshops (averaging 30 contact hours) was held for junior staff in the line ministries on the subjects of statistics, computer literacy, research methodology, and other techniques of policy analysis. The goal was to improve the ability of these staff members to report basic statistics on services delivered in their ministry. Enrollment was large, but few people took more than one course, and it is not clear that they use the training on their jobs.

The SLC has continued to be conducted every year, with decreasing reliance on technical assistance. Recruitment of qualified computer analysts has been a problem, partly because of low public sector wages and partly because very few people are trained to work with statistical software. Seminars for policymakers were scheduled to integrate policy research into decisionmaking but were never implemented. This [illegible] policymakers' interest in obtaining data-driven answers to policy questions and in ensuring staff involvement in various other parts of the project.

The project was designed with funds to support an annual agenda of social research. The initial call for proposals was issued in the first year of the project. Two years after the start of the project, no research had been funded. Twelve proposals were approved in the last three years of the project, but several of these are of more [illegible] sophistication than originally hoped for. The poor showing on this component may reflect some shortcomings in the details of how the call for proposals and review process were done. More likely, it indicates the difficulty (exacerbated by the failure to implement the policymaker seminars) that policymakers and government analysts have in posing policy-relevant research questions and understanding how they could be answered with quantitative data. Another factor was probably limited experience with proposal preparation among researchers (which might have been addressed earlier in the university training courses).

In the 11 years since the first SLC, the capacity for monitoring [illegible] and the amount and quality of quantitative policy analysis done in Jama[illegible] gradually increased. As a result the nature of the public policy debate ab[illegible] cial issues has changed somewhat. Capacity has yet to reach the level of sop[illegible] tion originally envisioned, however. The government has expressed its in[illegible] continuing to strengthen policy analysis capacity, but external fundin[illegible] required.

The Jamaican experience demonstrates that local capacity to implement [illegible] survey can be strengthened and that local analytic capacity can be built [illegible] a slow and incremental process. The challenge is to ensure that these gain[illegible] tainable, including the ability to produce a continuous supply of trained [illegible] ongoing demand for policy analysis by decisionmakers.

Lessons Learned

Experience in the five countries selected indicates that integrating [illegible] analytic capacity with data collection efforts provides an immediate [illegible] useful data source that can be used as a learning device for analysts and p[illegible] Although some of these lessons may appear obvious, experience shows [illegible] often been ignored and are thus worth highlighting. We break the lesso[illegible] groups: four requirements for building capacity—explicit planning [illegible] open data access policy, and demand for policy analysis—and four [illegible] capacity—training, technical assistance, participation in research [illegible] establishment of policy analysis institutes.

Requirement One: Explicit Planning

Real and sustainable capacity building is unlikely to occur in the abse[illegible] articulated goals and strategies. Clarity is required from the outset [illegible] extent of existing in-country policy analysis capability, the types of [illegible] built, and the plan for doing so. Planning for capacity building [illegible] project planners make decisions about:

- The agency or agencies to be responsible for data collection, data pr[illegible] analysis (initially and over time), and data dissemination.
- The aspect(s) of capacity to be strengthened. For example, capacity [illegible] focus on data collection for production of the statistical abstract, on [illegible] within sector ministries to utilize survey data, operational statistics, an[illegible] for planning purposes, or on the capacity of public sector and universit[illegible] and private consultants to undertake more complex analyses using [illegible]

tabulations but not enough to pose policy questions or bring to bear more sophi[illegible] cated analytic techniques (as the Vietnam experience illustrated). Single round [illegible] veys can also help to build awareness of the need for a later, more extensive [illegible] and may serve as a pilot to explore particularly sensitive issues, assess risks, and [illegible] tune objectives and plans (as demonstrated in all five countries). While single [illegible] survey projects will be implemented from time to time, they can provide [illegible] modest aid in building analytic capacity.

Requirement Three: Open Data Access Policies

Efforts to build analytic capacity around a survey can not even begin unless [illegible] data are broadly accessible. Both the formal policies governing access to data [illegible] services for disseminating data sets affect the extent to which researchers [illegible] data. Open data access is required because the bulk of policy analysis, and [illegible] of activities to build it, must take place outside the central statistics [illegible] policy analysis requires detailed knowledge of current and proposed [illegible] programs and often of advanced statistical techniques — skills that are [illegible] and butter of survey institutions. Moreover, statistical agencies do not [illegible] Their ability to conduct policy analysis is further limited in that they [illegible] only a handful of staff engaged in analysis, and they are usually subject to [illegible] plan new surveys. Actors outside the statistical agency must therefore be [illegible] programs to build analytic capacity. This, in turn, requires an open data access [illegible] The capacity-building program should, however, include the staff of the [illegible] agencies to help them be more aware of their clients' concerns and to remain [illegible] partners in the efforts to provide good data for policy formulation and [illegible]

The importance of access to data was somewhat hidden in the presentation [illegible] individual case studies because, in fact, all five countries instituted open [illegible] policies and included participants from a range of government agencies [illegible] the countries organized data dissemination services. South Africa and [illegible] vided copies of the data sets to the participants in the workshops and [illegible] nars. Open access to LSMS data does not necessarily imply open access [illegible] sets. The government of Vietnam, for example, maintains open access [illegible] data but continues to deny researchers access to other data sets.

Requirement Four: Demand for Policy Analysis

Policy analysis should be driven by the needs of policymakers. The [illegible] policymakers in developing countries frequently do not know how to frame [illegible] questions or use the results of policy analysis. A common theme in the [illegible] the need to convince government officials that data can enhance public policy [illegible] sions. Making sure from the outset that policymakers are included in the [illegible]

ivities to build analytic capacity is vital. For example, a presentation of the results of n a single round of a survey can be a useful vehicle for generating excitement regard- the potential contribution of such analysis to policy decisions. At the dissemination rkshop in the Kyrgyz Republic, policymakers were invited to request custom tabu- ions that were conducted the very next day. In South Africa the dissemination work- ops were tailored to interest policymakers in the results of work that had been done the South African data and additional examples of work in other countries that uld be replicated. In Vietnam analysts participated in the training along with their n computer technicians so that the policy analysts could pose questions for which programmers would often generate answers. One-day workshops in each ministry uld also be useful in orienting individuals to data-driven policy analysis. Disseminat- the survey results, including the statistical abstract and working papers, can also mulate the interest of program managers in data-driven analysis. But it is important remember that one-day workshops do not generate sustained demand; continued orts are necessary to convince decisionmakers of the gains to be expected from in- ting in such projects.

deally there is a virtuous circle among policymakers, analysts, and data collec- s. Analysts will, unbidden, produce analysis useful to understanding good policy. cymakers will be influenced by the results and will commission work on specific es on alternative policies. Both efforts will drive good data collection. Initiating circle requires breaking into the loop on both the demand (policymaker) and ply (policy analyst and data collector) sides.

nd One Training

ritical mass of trained policy analysts is required before significant changes in the cy analysis environment can be effected. Because sending large numbers of indi- uals abroad for training is neither feasible nor desirable, creative approaches to in- untry training should incorporate a mix of short- and long-term studies.

Policy analysis encompasses a range of techniques, from simple descriptive tabula- to complex statistical analysis. Most analysts, especially those in sector minis- need to be able to analyze a mix of operational statistics as well as budget, and other data for monitoring and for planning. A smaller number, includ- senior research staff in government and researchers in universities and other re- h centers, need advanced analytic skills to answer more complex research ques- about the efficiency and effectiveness of programs. Training will often be needed th levels.

Evidence from the countries studied indicates that short-term training—probably st six weeks in duration—can help to improve a nation's capacity for basic analysis including the preparation of statistical abstracts and simple two-way tables, but can be relatively effective in enabling researchers to produce useful output from

raw data. The ability to formulate research questions, to undertake more com[illegible] analysis, and to prepare research reports requires more extensive formal tr[illegible] however. The inclusion of such training as a component of several projects has [illegible] vided a wider perspective for assessing its effectiveness. The first lesson is th[illegible] demand for training is likely to be high. Second, selection procedures for tr[illegible] programs should be standardized to ensure that all participants have the ne[illegible] prerequisites for participation, including some knowledge of computers and [illegible] to a computer outside of class. Third, participation should be limited to staff wh[illegible] jobs require such training. Finally, participants should be required to enroll [illegible] comprehensive upgrading program rather than in individual courses. These me[illegible] nisms will reduce class size, permit more interactive teaching approaches, and weed out less committed participants. Judging from experience in Jamaica and [illegible] nam, refresher courses in mathematics and computers may be required [illegible] participants for higher-level studies.

An important lesson highlighted by the Vietnam study is that trainees [illegible] have opportunities to interact with experienced analysts. Very few participants [illegible] this study used their newly acquired skills when they returned to work. Skills [illegible] likely to atrophy when this happens.

One-shot training programs, no matter what the length or the level, do not [illegible] future needs for trained analysts. Sustainable capacity building will require [illegible] ing graduate education to ensure the availability of ongoing training in policy [illegible] including training in a range of sophisticated research and statistical techniques. [illegible] relatively easy and fruitful approach is to encourage the use of the survey data [illegible] ever undergraduate or graduate statistical or policy analysis courses exist at [illegible] universities (through an open data access policy, well documented data sets, [illegible] haps the funding of special pilot projects or a data bank). Even this requires [illegible] initial analytic and processing capacity, however. Assembling a critical mass of [illegible] to teach sophisticated research and statistics courses may be difficult. [illegible] ing faculty is a long-term effort. At the same time, it may be difficult to re[illegible] trained faculty. Clearly, fortifying or establishing graduate programs will [illegible] sible in every country. Regional approaches to capacity building, such as the [illegible] Economic Research Consortium, which is attempting to strengthen grad[illegible] and research in Sub-Saharan Africa, present a viable alternative (Fine 199[illegible]). [illegible] tended benefit of the Jamaica project is that the University of the West Indies [illegible] maica is a regional institution and non-Jamaican graduate students are en[illegible] upgrading program.

Tool Two: Technical Assistance

The case studies demonstrate the importance of technical assistance for many [illegible] of the LSMS process. They show that the programs are most effective when [illegible]

fighting, and interagency diplomacy may mean that the administrator is unabl[illegible] devote significant time to the substance of research.

The sustainability of such efforts frequently depends on the level of political s[illegible] port they enjoy. A minimal level of support will be required even to formula[illegible] capacity-building effort, but the project itself can generate further demand for d[illegible] driven analysis. A household survey is a useful mechanism for sensitizing polic[illegible] to the types of questions that can be answered with empirical data and to stim[illegible] interest in quantitative analysis of policy questions.

Notes

Lorraine Blank is a consultant at the World Bank, and Margaret Grosh is a senior [illegible] Human Development Network at the World Bank. This paper relies greatly on five case [illegible] the many people who contributed to them.

• The Bolivia case study was written by Manuel Contreras and benefited from comm[illegible] Behrman, Christian Darras, Vicente Fretes Cibils, George Gray Molina, Menno Pradhan [illegible] Requena, Eugene Szepesy, Rosa Talavera, and Nico van Niekerk.

• The Jamaica case study was written by Lorraine Blank, Margaret Grosh [illegible] with comments provided by Pat Anderson, Paul Glewwe, P. B. K. Murthy, Jacques [illegible] and Colin Williams.

• The Kyrgyz Republic study was written by Raylynn Oliver, with comments provided [illegible] Falkingham, Valentina Lomakova, Michael Mills, and Alexey Proskuryakov.

• The South Africa case study was written by Francis Wilson and Dudley Horner [illegible] by Harold Alderman, Ann Duncan, Stephen Klassen, Carlo del Ninno, and Neeta Sirur.

• The Vietnam case was written by Lisa Drummond, with comments by Sara Bales [illegible] Paul Glewwe, Dominique Haughton, Jonathon Haughton, Jennie Litvack, Adam Mc[illegible] Prescott, Jens Wandel, and Diep Vuong.

• The synthesis paper benefited from comments by Ann Duncan, Paul Glewwe, [illegible] Haughton, Emmanuel Jimenez, Jennie Litvack, John Newman, Raylynn Oliver, [illegible] Neeta Sirur, and participants in an informal seminar in the World Bank. It was edited [illegible] Mackintosh. The Development Research Group of the World Bank provided fund[illegible] consultancies involved in all aspects of the study.

1. The survey was carried out under the direction of researchers from the University [illegible] lina, Paragon Research International, and the Institute of Sociology of the Russian Academ[illegible]

2. Modeled on the Detroit Area Study.

3. In Vietnam universities are teaching institutions, mainly at the undergraduate [illegible] mally do not have research programs.

4. In previous surveys data were collected and aggregated at the district level. [illegible] were forwarded to the provincial level, where they were again aggregated and sent [illegible] national-level aggregations were compiled. There was no file that contained the original [illegible] for each household.

References

The word "processed" describes informally reproduced works that may not be commonly [illegible] through library systems.

Managing Government Exposure to Private Infrastructure Risks

Timothy Irwin • Michael Klein • Guillermo E. Perry • Mateen Thobani

[illegible] privatization of infrastructure should lead to the development of new infrastructure, [illegible] ments in the operation of existing infrastructure, and a reduction in budgetary [illegible]. Whether countries reap the full benefits of privatization, however, depends on [illegible] are allocated. If, as is often the case in developing countries, governments assume [illegible] should be borne by investors, they may reduce incentives for efficiency and incur [illegible] liabilities. To solve these problems, governments need to improve their policies [illegible] their risk bearing to certain political and regulatory risks over which they [illegible] control. When a government provides guarantees, it should attempt to mea[illegible] cost and improve the way they are handled in the accounts and budgets. Mea[illegible] and budgeting are critical to improving decisions about the provision of guar[illegible] to improving project selection and contract design, and to protecting governments [illegible] entering into commitments that might jeopardize future budgets.

[illegible] gence of private infrastructure in the past decade should have large benefits [illegible] developing countries. Compared with government enterprises, private firms typi[illegible] stronger incentives to build and run infrastructure businesses effectively [illegible] low cost. If prices reflect costs and the firm's profits depend on consumer [illegible] private firms tend to choose good projects. Privatization also encourages [illegible] the imposition of cost-covering tariffs, thus addressing the problem of [illegible] ing that has afflicted many publicly provided infrastructure services. Greater [illegible] and cost-covering prices allow firms to make investments and provide ser[illegible] might not otherwise have been possible. They simultaneously improve the [illegible] fiscal position by making available the same quantity and quality of [illegible] with smaller budgetary subsidies.

[illegible] structure subjects private investors to major risks because the investments are [illegible] large and their costs can be recouped only over long periods of time. Two [illegible] features of infrastructure create additional risks. First, the investments are [illegible] sunk; the assets cannot be used elsewhere except at great cost. Second, infra-

structure projects often provide services that are considered essential and are p
vided by monopolists. As a result, services are highly politicized. This combinat
of factors makes investors especially vulnerable to opportunistic government actio

Before the investment is made, the government has every reason to promise
treat the investor fairly—to allow cost-covering tariffs and to avoid changing regu
tions in ways that would adversely affect the investor. Once the investment is m
however, the government has an incentive to renege on its promises. The gove
ment can satisfy political demands to reduce prices or otherwise appropriate
investor's profits without causing the investor to pack up and leave. Lax and Sebe
(1981) discuss the problem in the context of mining. These incentives make retur
to private investors uncertain and more sensitive to the host government's behav

To protect themselves from such risks as nonpayment by purchasers, cost o
runs, and low demand, private investors often ask the host government to pro
extensive guarantees. In other words, they ask the government to enter into so
form of arrangement that results in the net wealth of the government — not the p
vate investors—varying with the risky outcome. Eager to encourage investment w
out making any cash outlays, governments often consent.

Poorly designed guarantees threaten to undermine the benefits of privatiza
First, they can blunt the private investors' incentives to choose only good pro
and to run them efficiently. If the government bears the risk of the project's fail
the private investor will invest in projects that are more likely to fail; having inve
in a project, the private investor has little interest in maximizing its chance of s
cess. Second, guarantees may impose excessive costs on the host country's taxp
or consumers and expose them to too much risk. Because guarantees rarely sho
in the government's accounts or budgets, governments may not know the exte
their exposure. Moreover, economic crises can trigger many guarantees simultane
many of the government's contingent liabilities can thus become actual and cur
all at once. At worst, the issuance of guarantees may contribute to crises by en
aging excessive entrepreneurial risk taking (Merton 1978). Guarantees also may
to asset stripping, in which a firm's insiders extract value from the firm even
drive it into bankruptcy or excessive foreign borrowing (Akerlof and Rom
McKinnon and Pill 1995).

Policy Reforms to Reduce Risks

Governments issue guarantees to make projects attractive to investors, often
risk bearing as a way to compensate for shortcomings in current and expected
policies. One of the best things governments can do to make projects more attra
without issuing guarantees is to put in place good policies that generally reduce risk
and raise expected returns. Stable macroeconomic policies, for example, reduce

elihood of large changes in exchange and interest rates and therefore lessen the ssures on governments to prevent convertibility or transferability. The regular closure of timely and reliable information on the state of the economy and the ernment's finances makes it easier for investors to forecast future revenues. Lib- capital markets permit investors to spread risks more broadly—both locally and rnationally—and allocate them to those most willing to bear them. Good regu- ry policies reduce the risk that investors will be exploited after they have invested. ating nonpolitical regulatory agencies, for example, reduces investors' fears that ticians will keep service prices too low to allow an adequate rate of return. Strength- ng the independence and quality of the judiciary reduces investors' fears of being reated by the executive. Permitting international arbitration reduces investors' that they will be mistreated by local courts that are not independent.

ot surprisingly, governments that have established good policies and persuaded stors that those policies will be maintained can attract private investment with- tensive risk bearing. In the United Kingdom, for example, the government cts large amounts of private investment despite its policy of not bearing even ators risks except where they relate specifically to a project (United Kingdom . When developing countries have introduced good policies and maintained for a few years, they have also been successful in attracting private infrastruc- apital without guarantees. In Argentina, for example, the complete restructur- nd privatization of the power industry has permitted the government to attract te investment without having to assume major risks or issue guarantees (Klein . In Chile private firms recently have invested in telecommunications, power, s without government guarantees (Jadresic 1997).

this article, we focus not on the surrounding policy environment but on the priate allocation of risks of private infrastructure projects among governments, investors, and consumers. We also discuss how governments should budget count for risks.

s of Risk

ments, consumers, and private investors bear various types of risk, including de- and payment risk, exchange and interest rate risk, and political and regulatory s also bear implicit risk and other types of risk. See box 1 for more details on g risk

nt and Payment Risks

tizing toll roads, the host government often commits itself to ensuring that ate owner receives at least a minimum level of revenue when demand is lower

Box 1. ***Defining Risk***

In theoretical treatments of finance and government risk bearing, as well as in the applied fi[illegible] securities analysis and portfolio management, "risk" is often used to refer to the volatility [illegible] around an average or expected return (see, for example, Markowitz 1991 and Arrow and [illegible] In this sense, risk is equivalent to the statistical concept of variance, and a project's risk can [illegible] without any change in the expected (or mean) return on the project. Investors who were [illegible] (in the sense used in economics and finance) would be indifferent to risk in this sense, and risk [illegible] sort can be effectively eliminated by diversification if it is not systematic.

By contrast, in project finance, "risk" frequently refers to the ways in which actual result[illegible] worse than planned. Here the benchmark is not the expected return of the project but th[illegible] higher) return that investors would receive if everything went according to plan. For exampl[illegible] estimate the returns they will earn on the assumption that the government will not expro[illegible] investment, while noting the risk of expropriation. An increase in expropriation risk [illegible] not just increase the volatility of returns, it reduces the expected return. Even risk neutra[illegible] would prefer to avoid these "risks." Diversification cannot eliminate this risk; it can only [illegible] loss among many people.

We use "risk" in the sense of variance or volatility around a statistically expe[illegible] Expropriation risk, for example, is thus the volatility in returns around an expected return [illegible] to uncertainty over whether the government will expropriate.

than expected. This commitment shifts some of the risk of variation in de[illegible] the government. In the El Cortijo–El Vino toll road project in Colombia [illegible] ample, the government undertook to reimburse the concessionaire if traffi[illegible] than 90 percent of the specified level. The government agreed to pay the con[illegible] aire an amount equal to the toll times the difference between 90 percent [illegible] estimated number of vehicles and the actual number of vehicles (Lewis and [illegible] 1997).

Governments bear similar risks in other sectors. The Colombian go[illegible] instance, provided a minimum revenue guarantee when it awarded a [illegible] transfer concession for a new runway at Bogota's El Dorado airport [illegible] 1996). And many governments, through their utilities, have agreed to p[illegible] dent power producers a fixed amount each year that is independent of [illegible] level of power subsequently demanded from them.

An agreement by a state-owned utility to pay an independent powe[illegible] irrespective of demand protects the investor from the risk of falling de[illegible] power or of new and cheaper generators coming on stream in the future. [illegible] not protect the investor from the risk of the utility defaulting on its obliga[illegible] protect themselves against this risk, investors usually ask the governmen[illegible] more creditworthy than the utility it owns, to guarantee the utility's payme[illegible] some cases, as when Pakistan sought to expand its power generation capacit[illegible] tors ask their governments to ask multilateral agencies to guarantee payment

Exchange and Interest Rate Risks

Governments have sometimes borne the risks associated with adverse fluctuations in exchange and interest rates. The Spanish government, for example, had many private toll roads built during the 1960s and early 1970s and bore the exchange rate risk on the foreign loans that financed the roads. Gómez Ibáñez and Meyer (1993:126) describe the government guarantees and their rationale:

> The Spanish government had required the early concessions to finance a large part of their costs from foreign debt in order to ease Spain's balance-of-payments problems and to avoid drawing away domestic savings from other projects. The 1972 law [on toll road concessions] set standards that at least 45 percent of construction costs be financed from foreign loans, at least 10 percent from equity, and no more than 45 percent from domestic loans. The early Spanish highway companies had trouble raising funds from foreign capital markets, however, and in return the government agreed to guarantee some of these loans and to protect the companies from exchange rate fluctuations. The 1972 law specified that the government would guarantee up to 75 percent of the foreign loans; moreover, all foreign loans would be denominated in pesetas, with the government assuming the full exchange rate risk.

[illegible] peseta depreciated relative to the foreign currencies in which the loans were [illegible] concessionaire's loan repayments would remain the same, but the Spanish [illegible] would make an additional payment to ensure that the foreign lenders [illegible] foreign currency [illegible] Spanish taxpayers spent about $2.7 [illegible] result of the guarantees.

[illegible] and Regulatory Risks

[illegible] often bear certain political and regulatory risks, even when they bear [illegible] risks mentioned above. In the Melbourne City Link, a private toll road [illegible] private sector parties bear most of the demand, payment, and exchange [illegible] rate risks. The state government bears several risks that are tied to ac[illegible] influenced by the government. For example, if the government subse[illegible] toll roads or takes actions that deliberately reduce the profitability of [illegible] investor, the government will compensate the investor. The government [illegible] the risk associated with possible court findings that aboriginal land rights have [illegible]. And it bears the risk that workers on the construction site will strike to [illegible] against the state government rather than as part of site-specific disputes. In [illegible] example, the government of Pakistan has a policy framework for private [illegible] generation. The government agreed to "cover certain political and govern-

mental force majeure risks, provide protection against changes in certain ta[illegible] duties, and ensure foreign exchange convertibility for the projects" (Internati[illegible] Finance Corporation 1996:49).

Implicit and Other Forms of Risk Bearing

Government risk bearing need not be made explicit in contracts or laws. Someti[illegible] everyone concerned expects that the government will in fact bail out a company [illegible] would otherwise fail. The case of private Mexican toll roads may provide such [illegible] example. They were partly financed by commercial banks, which were owned [illegible] time by the government. Some observers have argued that the commercial b[illegible] exercised less care than they should have in assessing the credit risks they assum[illegible] Although the Mexican government did not explicitly agree to bear the credit r[illegible] taken on by the banks, it did in the end bail them out when they got into tro[illegible] partly as a result of the poor financial performance of the private toll roads. [illegible] argue that the banks had expected the bailout and that this expectation had an e[illegible] similar to that of an explicit government guarantee.

Governments bear risk in other, less obvious ways as well. They may lend di[illegible] to projects and bear repayment and perhaps interest rate risks. They may be part owners of a project and thereby bear a proportion of the overall risk [illegible] project. Moreover, governments own a share of many firms, in an economic if [illegible] legal sense, through the corporate tax system: if profits are high, the governmen[illegible] more corporate income tax; if they are low, it gets less.

Principles of Risk Allocation

Infrastructure project risk can be allocated, at a broad level, to governments, [illegible] or consumers. How should governments decide whether to bear risks in [illegible] infrastructure project? If they do decide to bear risk, which risks should they [illegible]

Two critical factors determine whether an agent should bear risk: the deg[illegible] which the agent can influence or control the outcome that is risky and the [illegible] ability to bear the risk. These two factors often push in different directions [illegible] group or organization that has most control over the risky outcome may not [illegible] the best position to bear the risk. Other things equal, risks should be allocated [illegible] agents who can best control the risky outcome and to agents who can bear the [illegible] the lowest cost. Those agents are the least risk averse because they can most [illegible] insure or hedge against the risk, or because they can spread the risk among [illegible] people. The two critical risk factors must be balanced by three other factor[illegible] incentives to reduce risk, the transactions cost of allocating risks, and second[illegible] considerations stemming from credibility issues and policy transitions.

[illegible]ontrol over Risk and the Cost of Bearing Risk

[illegible]he benefits of allocating risks to those who can best control them must sometimes [illegible] weighed against the benefits of allocating risks to those who can bear them at least [illegible]. Spreading a risk among many shareholders or taxpayers may lower the costs of [illegible] bearing, but allocating a risk to a small number of agents who have control over [illegible] risk may help ensure the success of the project. Rather than diversifying risks [illegible]mpletely, then, it usually pays to give managers and strategic investors significant [illegible]kes in the project. A single risk-averse entrepreneur would face a higher cost of [illegible] demand risk than a government, but the entrepreneur would face stronger [illegible] to increase demand and reduce overall risk.

[illegible] telecommunications concession, for example, the demand risk would be the [illegible] demand may be higher or lower than forecast. The principle of control over [illegible] suggests that the firm should bear this risk because it can effectively [illegible] demand by keeping service quality high. By contrast, convertibility risk— [illegible] that local currency may not be convertible into foreign currency—should be [illegible] by the government because it alone controls this risk.

[illegible] the telecommunications example, the government of a large country with [illegible] can bear demand risk at lower cost than a small telecommunications [illegible] large telecommunications firm with lots of small and diversified share- [illegible] assume demand risk at a lower cost than the government of a small [illegible] correlation of [illegible] other assets of the agent bearing the risk also [illegible] not be desirable for the government to bear demand risk because the [illegible] likely to suffer the consequences of low demand just when its tax [illegible]. Foreign investors may be better placed to assume this risk be- [illegible] hold a portfolio of assets whose value is not correlated with local [illegible] conditions.

[illegible] Incentives, Transactions Costs, and Second-Best Considerations

[illegible] government [illegible] good macroeconomic, legal, and regula- [illegible] firms may be reluctant to invest without government guarantees be- [illegible] question whether the reforms will be maintained or protected by courts. [illegible] country's laws and regulations are as good on paper as those of, say, [illegible] States, investors will be considerably more wary about investing there. In [illegible] governments may have to provide explicit undertakings—to allow con- [illegible] to compensate in case of expropriation, for example. In countries such [illegible] States, investors may think it unnecessary to seek certain explicit project- [illegible] guarantees, either because the risks are negligible or because the investors are [illegible] that the legal system and the courts will protect them in case of problems. [illegible] that have reformed their policies only recently, investors may want the

government to assume these risks explicitly in a contract, often one that is guaranteed by a multilateral agency.

Allocating risks that are better controlled by the government to the government will improve outcomes only if the government responds to financial incentives. [illegible]ernments are generally less responsive than firms to financial incentives because [illegible]ernment decisionmakers often do not act in the interests of the citizens. W[illegible] financial consequences of the government's risk bearing do not show up in [illegible]gets or accounts, government may be less responsive still. If, for instance, the [illegible]hood of the government's permitting convertibility is unaffected by [illegible] to pay compensation in case of malfeasance, there is no value in allocating this [illegible] government. Similarly, as decisionmakers face weak incentives to low[illegible]ernments may be less adept than private investors at taking advantage [illegible]ties to reduce risk, for example, through diversification or hedging.

Transactions costs should also be considered. An allocation assigning a project's many risks according to each party's control over the outcome [illegible] of risk bearing may require detailed analyses, tough negotiations, compl[illegible] tracts, expensive monitoring arrangements, and possibly the high cost [illegible] disputes in court. The optimal allocation of risk takes these costs into [illegible]

Second-best considerations may also come into play. According to the risk allocation discussed here, many governments in the developing world [illegible] bear too much risk when they privatize infrastructure. This situation [illegible] still be better than the alternative of public ownership. When the go[illegible] all the commercial risk, privatization almost always transfers some risk [illegible] sector. When the government guarantees the sponsor of [illegible] percent of expected toll revenue, for example, the government bears [illegible] would if it owned the road and bore 100 percent of the risk. [illegible] road and providing guarantees to the concessionaire may be better [illegible] government build and operate it without private participation or [illegible] at all. Similarly, governments that are unable to carry out all the [illegible] attract private infrastructure investment without government guarantees [illegible] the second-best option of privatization without the full transfer of [illegible] to private investors. In many such cases, however, a direct fiscal [illegible] preferable to a government guarantee.

Allocation of Typical Infrastructure Risks

Typical risks that governments are asked to assume in infrastructure [illegible] projects include political, regulatory, quasi commercial, demand and const[illegible] cost, and exchange and interest rate risks. This section provides some practi[illegible] lines on how to allocate these risks.

little with increases in quality. Therefore, governments need to balance their [illegible] to provide incentives for proper project screening with the recognition that [illegible] tors often have limited control over demand.

By changing the way they regulate infrastructure, governments can reduce [illegible] mand risk faced by concessionaires and thereby reduce the concessionaires' dem[illegible] guarantees. Instead of auctioning the right to operate the service for a fixed per[illegible] time, as is typical in most road and bridge concessions, the term of the oper[illegible] cession could be made to vary with demand. If demand is higher than expect[illegible] concession will be shorter; if demand is lower, the concession will be longer [illegible] method, which has been used in the United Kingdom for bridges, reduces risk

In an ingenious variant of this method, Engel, Fischer, and Galetovic (199[illegible]) [illegible] pose to award the concession to the bidder seeking the lowest present value [illegible] enue, calculated with a discount rate specified in advance by the government [illegible] concession ends when the concessionaire's revenue reaches the present value [illegible] sought. The concessionaire still bears some demand risk—if demand is to[illegible] revenue may never reach the target value—but it bears much less. Moreov[illegible] investor still has an incentive to select only those projects that are likely to be [illegible] cially attractive without government subsidies

Floating-rate loans fund many infrastructure investments, making the pr[illegible] profits highly sensitive to changes in interest rates. Projects often involve co[illegible] able foreign financing. If project revenues are in local currency and the inv[illegible] want to earn foreign-currency profits, foreign investors will suffer if the lo[illegible] rency depreciates. Ensuring that the right parties bear *exchange and interest ra[illegible]* thus important to the success of the project. Should the government accede t[illegible] tor demands to assume exchange and interest rate risks

At first glance, our framework appears to suggest that the government should [illegible] infrastructure risks because it can better control them. Private investors have [illegible] no control over the exchange rate or prevailing interest rates. If governments [illegible] such risks, they have a financial incentive to adopt macroeconomic policies [illegible] to prevent depreciation or increases in interest rates. Such policies may [illegible] useful signaling effect. Governments that plan to adopt imprudent mac[illegible] policies will find risk bearing expensive. Most likely, governments th[illegible] issue such guarantees will act reasonably (at least if politicians are conc[illegible] the government's fiscal position)

Mas (1997) argues cogently, however, that three other factors often [illegible] incentive and signaling benefits. First, it is difficult to separate the effect [illegible] profitability of exchange or interest rates and business decisions. Losses [illegible] rency depreciation, for example, could be blamed on the government, which [illegible] the currency to fall, or the firm, which left itself exposed by borrowing in [illegible] currencies. Second, in flexible exchange rate regimes, exchange rate guarantees [illegible] have undesirable as well as desirable incentive effects on the government [illegible]

will make the full payment), the government's expected cost of the guaran[illegible] $100,000. For more realistic cases, the expected cost may be more difficult to cal[illegible] late. There may be more than two relevant possibilities, and the estimation of probabilities may be extremely difficult.

Nevertheless, the calculation of expected losses is sometimes feasible using [illegible] tively straightforward techniques. The most tractable cases are those in which [illegible] government has issued a large number of similar guarantees for many years and recorded information on defaults. In such cases analysts can calculate the expe[illegible] cost of the guarantees in the same way as, say, car insurance premiums are cal[illegible] The reforms the U.S. government enacted with the Federal Credit Reform [illegible] 1990 are examples of this (Lewis and Mody 1997).

The cost of some unique guarantees can also be estimated simply. [illegible] guarantees, in which the government guarantees the repayment of a loan [illegible] other party can, for example, be calculated by looking at the interest rate [illegible] on guaranteed and nonguaranteed loans. If a firm pays 15 percent interest [illegible] nonguaranteed debt and 10 percent on loans guaranteed by the government [illegible] annual value of the guarantee is 5 percent of the amount borrowed [illegible] Patro 1995).

Analysts can value guarantees and contingent liabilities, including more [illegible] cated ones, using the techniques developed in the past 25 years to [illegible] derivatives (such as options, futures, and swaps). Extending a credit guarantee, [illegible] example, is equivalent to the government's selling a put option to the lender [illegible] gives the lender the right to put the loan to the government. The valuation [illegible] guarantees requires the skills of financial specialists, however, and the [illegible] timely, reliable, and cost-effective valuation has not yet been widely tested. [illegible] possibilities are not merely theoretical: guarantees in both Colombia and the [illegible] States have already been valued using option pricing techniques (Lewis and [illegible] 1997).

Valuing the government's guarantees and other contingent liabilities [illegible] simply noting maximum exposure—has important advantages. By [illegible] expected cost of the government's guarantees, analysts can more easily [illegible] antees with cash subsidies. When guarantees are not valued, a government [illegible] fer to provide a guarantee instead of a subsidy even if the guarantee costs [illegible] the subsidy, because a future administration may bear the hidden costs of [illegible] antee. When guarantees are valued, policymakers are more likely to make [illegible] on the basis of real rather than apparent costs and benefits.

Incorporating Expected Losses in Accounts and Budgets

If analysts can reliably calculate expected losses, the government should incorp[illegible] the estimates in its accounts and budgets. Most governments have cash-based [illegible]

that its daily value at risk at the 1 percent level is $10 million, meaning that the
only a 1 percent chance, under normal market conditions, that it will lose more t
$10 million in the next day (Jorion 1997). Although banks were the first organ
tions to report value at risk, the principles behind such reporting apply to all org
zations. Governments should think about how they can apply these principle
their own operations and what useful estimates they might be able to publish

Taking a Government-wide Approach to Risk Measurement

Expected losses can be measured individually and then aggregated; the governme
total expected loss from issuing several guarantees is simply the sum of the expe
losses associated with each guarantee. Normally, risks cannot be estimated indiv
ally and then summed because the total risk depends on the relationships betw
the individual risks.

A government worries not about the risk relating to any one guarantee but
riskiness of its portfolio of assets and liabilities. The reporting of risk exposure
most useful when done for the government as a whole. Therefore, measuring
true risks associated with infrastructure privatization requires an assessment of
riskiness of all of the government's operations. Exchange rate guarantees
instance, are likely to be more risky if the government also has net debt denomin
in foreign currencies. A depreciation of the local currency may simultaneo
increase debt service payments and trigger payments under the guarante
portfolio-wide approach also allows the government to judge the importan
the risks associated with infrastructure privatization relative to the risks asso
with government debt, pensions, the banking system, and debt owed by its
provincial governments.

The creation of consolidated accrual accounts—incorporating present values wh
practicable—is a step toward portfolio-wide risk monitoring. Combined with
information on likely future revenues and expenditures, the balance sheet can
vide an indication of the government's vulnerability to exchange rate, interest
and other shocks. It can indicate the relative importance of monitoring and m
ing infrastructure guarantees on the one hand and debt on the other. Ultim
governments should aim to measure portfolio-wide value at risk. The achieve
of that goal appears to be a long way off, however.

Risk Management

A government is in a good position to consider managing its portfolio only when
has good information on the risks to which its total portfolio is exposed. It c
manage to reduce those risks, for example, by entering into derivative contracts
hedge against changes in exchange rates or commodity prices. Whether the gove

nent should act to reduce risk (rather than just reduce expected costs) is perhaps an open question. It is analogous to the firm's decision about whether to attempt to educe the variance of shareholders' returns or simply to maximize the expected value of the returns.

Citizens, like shareholders, are usually risk averse. The policy decision is whether he government should manage risk on their behalf or publicize its risk exposure and permit taxpayers to diversify and hedge their portfolios in ways that give them the isk exposure they want. In practice, citizens often lack sophistication in considering isk. Moreover, because of transactions costs or government restrictions, citizens may have limited opportunities to hedge against government risk. In that case, governmental risk management may be desirable. It would aim to achieve a level of risk the government finds acceptable, given its citizens' risk preferences, at the lowest possible cost.[2]

Conclusion

Whether infrastructure privatization will realize its potential depends on how government allocates the risks facing privatized businesses. Government can increase the benefits of privatization by assuming risks it can control itself (convertibility risk, for example), but it should normally avoid bearing other risks. That way, investors face strong incentives to select projects well and to run those that they do select efficiently.

In many infrastructure privatizations, governments have assumed risks that investors should bear, both because the investors have been understandably wary of taking on the considerable risks involved and because governments have been able to offer guarantees without incurring any immediate cash costs. A government can thus take two steps to improve the environment for risk allocation. It can reduce the extent of the risks investors face by pursuing stable macroeconomic policies, disclosing information, implementing good laws and regulations, and liberalizing financial markets. And it can improve the way it measures, budgets, and accounts for the guarantees it does give, so that the costs and risks are clear at the time the guarantees are issued — not only when the government must subsequently pay up.

Notes

[illegible] is an economist in the Private Participation in Infrastructure section, World Bank; [illegible] is chief economist of Shell International; and Guillermo E. Perry is chief economist [illegible] Malcolm Thobani is principal economist in the Latin America and the Caribbean Region, World Bank. This paper summarizes papers and discussions of a conference held in Cartagena, Colombia, in [illegible] 1997 on the topic. The authors are grateful to conference participants for their contributions and special thanks to Katherine Brewer for her excellent organization.

1. For more on the analysis of value at risk, see Jorion (1997); J. P. Morgan's "RiskMetrics" documentation, available at *http://www.riskmetrics.com*; and the various reports available at *http://www.contingencyanalysis.com*.

2. In the extreme case of a guarantee for a risk that is negatively correlated with the value of the government's portfolio, assessment of risk in isolation would lead the government to think the guarantee created risk, when in fact it reduced it. In other words, because the value of this guarantee [illegible] to fall when the value of the rest of the government's portfolio rises, and vice versa, the total [illegible] of the government's wealth would be greater without the guarantee.

3. A useful short summary of risk management for countries (as opposed to governments [illegible] found in Claessens (1992). Claessens and Qian (1991) apply the techniques to African [illegible]

References

The word "processed" describes informally reproduced works that may not be commonly [illegible] through library systems.

Afterman, Allan B. 1997. *GAAP Practice Manual*. Vol. 1. New York: Warren, Gorham [illegible]

Akerlof, George A., and Paul M. Romer. 1993. "Looting: The Economic Underworld of Bankruptcy for Profit." *Brookings Papers on Economic Activity* 2: 1–73.

Arrow, K. J., and R. C. Lind. 1970. "Uncertainty and the Evaluation of Public Investment Decisions." *American Economic Review* 60: 364–78.

Claessens, Stijn. 1992. "How Can Developing Countries Hedge Their Bets?" *Finance and Development* 29: 13–15.

Claessens, Stijn, and Y. Qian. 1991. "Risk Management in Sub-Saharan Africa." In [illegible] and John Underwood, eds., *African External Finance in the 1990s*. Washington, D.C.: [illegible]

Engel, Eduardo, Ronald Fischer, and Alexander Galetovic. 199[illegible]. "Infrastructure [illegible] Government Guarantees." In Irwin and others, eds.

Gomez-Ibáñez, Jose A., and John R. Meyer. 1993. *Going Private: The International Experience with Transport Privatization*. Washington, D.C.: Brookings Institution.

International Finance Corporation. 1996. *Financing Private Infrastructure*. [illegible] Series 4. Washington, D.C.

Irwin, Timothy, Michael Klein, Guillermo E. Perry, and Mateen Thobani, eds. 1997. *Dealing with Public Risk in Private Infrastructure*. World Bank Latin American and Caribbean S[illegible] Washington, D.C.

Jadresic, Alejandro. 1997. "Comments on Government Support to Private Infrastructure in Emerging Markets." In Irwin and others, eds.

Jorion, Philippe. 1997. *Value at Risk: The New Benchmark for Controlling Market Risk*. [illegible] Irwin.

Juan, Ellis. 1996. "Privatizing Airports: Options and Case Studies." *Public Policy for the Private Sector*. *Special Edition, Infrastructure*. Washington, D.C.: World Bank.

Klein, Michael. 1997. "Managing Guarantee Programs in Support of Infrastructure [illegible]" Policy Research Working Paper 1812. World Bank, Private Sector Development [illegible] Washington, D.C.

Lax, David A., and James K. Sebenius. 1981. "Insecure Contracts and Resource Development." *Public Policy* 29(4):417–36.

Lewis, Christopher M., and Ashoka Mody. 1997. "The Management of Contingent [illegible] Risk Management Framework for National Governments." In Irwin and others, eds.

The Evolution of the World Bank's Land Policy: Principles, Experience, and Future Challenges

Klaus Deininger • Hans Binswanger

[illegible] article examines the evolution of policy recommendations concerning rural land [illegible] since the formulation of the World Bank's "Land Reform Policy Paper" in 1975. [illegible] paper set out three guiding principles: the desirability of owner operated family [illegible] the need for markets to permit land to be transferred to more productive users; and [illegible] importance of an egalitarian asset distribution. In the 25 years since that paper was [illegible] these guiding principles have remained the same, but it is now recognized that [illegible] tenure systems can be more cost effective than formal title, that titling pro[illegible] should be judged on their equity as well as efficiency, that the potential of land [illegible] markets has often been severely underestimated, that land sale markets enhance [illegible] only if they are integrated into a broader effort at developing rural factor mar[illegible] and that land reform is more likely to result in a reduction of poverty if it harnesses [illegible] than undermines the operation of land markets and is implemented in a decen[illegible] [illegible] land policies that incorporate these elements requires a coher[illegible] and institutional framework together with greater reliance on pilot programs to [illegible] applicability of interventions under local conditions.

[illegible] rural areas of most developing countries, land is not only the primary means for [illegible] a livelihood but often the main vehicle for investing, accumulating wealth, [illegible] transferring it between generations. Thus the ways in which access to land is regu[illegible] property rights are defined, and ownership conflicts are resolved has broad impli[illegible] beyond the sphere of agricultural production. These regulations, rights, and [illegible] affect not only the ability of households to produce for their subsistence [illegible] the market but also their social and economic status (and often their collective [illegible]), their incentive to work, their willingness to use the land sustainably, and [illegible] ability to self-insure or to obtain access to financial markets.

The importance of land issues in fostering economic growth and reducing [illegible] was the impetus for the World Bank's 1975 "Land Reform Policy Paper." [illegible] time, this dialogue was complicated both by an economic environment in [illegible] government interventions often caused the prices of rural land to deviate [illegible] cantly from the net present value of agricultural profits and by a political [illegible] which land was at the heart of a broader ideological struggle. In many [illegible] countries today, far-reaching macroeconomic reforms have removed [illegible] policies, the ideological divide has narrowed or disappeared, and the [illegible] structural issues has greatly increased the demand for policy advice. [illegible] ations provide an opportune moment to review earlier policy recommenda[illegible] to use experience to assess the role of such policies in the broader process [illegible] ment. This article reviews the analytical underpinning for policy [illegible] and examines the effectiveness of such advice in the areas of tenure [illegible] markets, and land reform

The broad consensus underlying current thinking about land issues [illegible] marized in four key principles:

- The desirability of owner operated family farms on both efficiency [illegible] grounds
- The importance of secure property rights to land [illegible] investment and in providing the basis for land transaction[illegible]
- The need for a policy and regulatory environment that promot[illegible] more efficient land uses
- The positive impact of an egalitarian asset distribution and the [illegible] tributive land reform where nonmarket forces have led to [illegible] ership and operational distribution of land [illegible] by very large and very small holdings

Although these principles remain valid, experience with land reform [illegible] earlier assumptions in four areas. First, the 1975 World Bank land [illegible] recommended that communal tenure systems be abandoned in [illegible] titles and the subdivision of the commons. Today it is recognized [illegible] munal tenure arrangements can increase tenure security and provi[illegible] basis for land transactions in ways that are more cost effective than [illegible] Where that is the case, governments may find it useful to reduce the [illegible] eration, improve accountability, and facilitate a gradual evolution [illegible] systems to meet emerging needs, possibly for greater individualization [illegible] rights over time.

Second, although individual titling has great potential to increase [illegible] productivity, several preconditions must be satisfied for this to be a [illegible] vention. The circumstances under which title is conferred are importa[illegible] ample, titling should be area based (that is, it should cover an entire [illegible]

and fit within a broader strategy of rural development. Otherwise, imperfections in other factor markets may undermine or even eliminate the advantages from possession of title, at least for the poor.

Third, the earlier skeptical view of land rental markets has given way to a recognition of their critical role as a means for providing the poor with access to land. The removal of remaining restrictions on land rental is therefore a top policy priority. In contrast, however, removing the restrictions on markets for land sales may not be the most urgent requirement for increasing efficiency and may have a negative impact on equity. Measures thus should be sequenced properly, emphasizing rentals rather than sales, and should be integrated with the development of other rural factor markets.

[illegible] fourth, a growing literature has made the case for redistributive land reform [illegible] efficiency and equity grounds. Most of the land reforms undertaken during the [illegible] 20 to 30 years, however, were politically motivated and have not lived up to [illegible] expectations. Recently, a new approach has emerged, encouraging community-[illegible]aged agrarian reform based on voluntary negotiation. Provided that careful moni-[illegible] permits officials to make the necessary changes in program design and that [illegible] pressures to provide free handouts to influential lobbies can be resisted, this [illegible] can help to overcome long-standing problems of asset distribution and [illegible] exclusion, which are key factors leading to rural violence.

[illegible] addition to such changes in specific recommendations, land policy is increas-[illegible] viewed as an integral element of a broader development process rather than as [illegible] of narrowly oriented technical interventions. This view is based on experi-[illegible] showing that a lack of consensus on the broader subject of land policy has often [illegible] the effect on development of specific interventions, such as land ti-[illegible] countries where land issues have in the past resulted in civil strife, revolu-[illegible] reaching a consensus requires time and involves all the relevant sectors [illegible]. Initiating such a process, ensuring its integration into a broader frame-[illegible] development initiatives, and strengthening the analytical capacity of [illegible] have become important components of the World Bank's approach to [illegible]. At the same time, the political sensitivity of such issues and the need to [illegible] specific conditions often dictates that specific approaches be explored [illegible] scale before they are implemented broadly.

[illegible] Conceptual Basis of the World Bank's Land Policy

[illegible] fact, confirmed by a large literature, is that owner-operated smallholder [illegible] desirable from both an equity and an efficiency perspective. Secure indi-[illegible] property rights to land would therefore not only increase the beneficiaries' [illegible] and provide collateral for further investment but, if all markets were com-[illegible] would automatically lead to socially and economically desirable land mar-

ments (Glover 1990). Economies of scale in processing or marketing are therefore important for the size of farming operations only so long as markets for outputs and inputs are either unavailable or malfunctioning.

Imperfections in other markets, such as lack of access to capital and insurance markets, put small farms at a disadvantage. Their limited ability to cope with [illegible] could offset the cost advantage that small family farms enjoy. For example [illegible] Sudan, capital market imperfections led to a positive relationship between farm [illegible] and productivity (Kevane 1996), while Burkina Faso recorded an inverse farm [illegible] productivity relationship (Udry 1996), and in other contexts an optimum farm [illegible] emerged (Carter and Mesbah 1993). To the degree that imperfections in these markets, rather than an inherent productivity advantage of large farms, are at the [illegible] of differences in the shadow price of land across categories of farm size [illegible] ments in these markets through regulation, better information, or cooper[illegible] reap economies of scale or input supply could lead to productivity gains

The Importance of Land Market Transactions

That well-functioning land markets can promote efficiency-enhancing [illegible] is well recognized. The extent of such land transfers is affected by government [illegible] cies, informational constraints, and incomplete credit markets and their [illegible] land prices, producers' ability to access financial markets, and transaction [illegible] ciated with land rentals and sales

Land Price Formation

Mortgaged land cannot be used as collateral for working capital. Thus the [illegible] would not reap the production credit advantage and would be unable to [illegible] from the income generated from the land. In addition, the value of the [illegible] unmortgaged land as collateral in formal credit markets would be capi[illegible] land prices, and the equilibrium price of land at given credit costs would [illegible] present discounted value of the income stream that can be produced from [illegible] (Binswanger and Elgin 1988, Just and Miranowski 1988). Credit subsidies [illegible] vantages, and the use of land as an inflation hedge would have the same [illegible] only income stream available to the poor for consumption is the imputed [illegible] family labor. The remainder of the profits would have to be used to pay for [illegible] reducing the purchaser's utility below what could be achieved in the labor [illegible] and implying that land purchases would normally have to be financed out of [illegible] hold savings. Even if credit access were perfect, mortgage based land acquisi[illegible] would be unlikely to lead to redistribution in favor of landless households (Binswanger and Elgin 1988). Any additional subsidies, tax advantages, and other [illegible]

ther increase land prices above the net present value of agricultural profits would, course, further reduce the scope for participation of the poor (Gunjal, Williams, d Romain 1996).

edit Market Imperfections

e same credit constraints that make it difficult for landless farmers to finance the rchase of land also affect their ability to pay for land in the rental market. Credit oning reduces their ability to make productive use of land, possibly outweighing e advantage of owner-operated farms, reducing their reservation price for land, d possibly leading to (relatively inefficient) rentals to large producers.

One outcome of these difficulties is that smallholders are forced to adopt costly nd relatively inefficient) insurance substitutes to enable them to deal with unex-cted productivity and idiosyncratic shocks, such as adjusting their crop and asset rtfolios to a low-return, low-risk combination that reduces their vulnerability. other possibility is to replace land with a more liquid asset such as grain, which en though less productive, would provide them with greater security in case of bsistence risk (Zimmerman and Carter 1999). Indeed, for India, Rosenzweig and olpin (1993) show that the lack of insurance substitutes affects farmers' invest-ents. Despite this less-than-optimal allocation of assets, small farmers are still more oductive than large farmers (Rosenzweig and Binswanger 1993). Dercon (1996) nd Dercon and Krishnan (1996) find that in Africa the levels of liquid assets and pital available to the household determine its ability to enter into high-risk but gh-return activities.

he market for land is also affected by credit market imperfections that deny allholders insurance against shocks, such as bad harvests or accidents, and may ce them to sell off land in periods of distress. Where the covariance of weather ks would imply wide fluctuations of land prices over time, it is hard or impossible r smallholders to recover from those asset losses because they would have to sell at w prices during disaster (when there is little effective demand) and buy during rmal times, when prices are high again (Bidinger and others 1991). In research on ngladesh and India, Cain (1981) illustrates this view that transactions in the land les market are driven by credit and insurance limitations rather than by cultivators' oductive efficiency. He finds that in villages that had access to a safety net pro-rge farmers generally sold land in response to shocks and to obtain cash for major estments such as drilling wells, purchasing pumps, and educating and marrying heir children. By contrast, where such consumption-smoothing devices were ab-sent the majority of sales were prompted by distress (to obtain food and medicine). Whether or not households were able to buffer consumption during crisis situations had a significant impact on whether markets helped to equalize or disequalize land endowments. Indeed, distress sales have not only played a major role historically but

also are linked in the literature to the elimination of traditional mechanisms fo
coping with risk (Kranton and Swamy 1997).

Transaction Costs and Incentive Issues

Because land is immobile, land markets are localized, with several important conse quences. As Balcazar (1990) and Carter and Zegarra (1995) note, land sales market are normally highly segmented, especially in countries with a dualistic distributio of ownership and where sales by large farmers to smallholders are virtually absent One explanation is that it is costly to subdivide large farms so as to make them suitable for smallholder cultivation. Similarly, as noted earlier, transaction costs fu ther discourage small land transactions or drive them into the informal marke

Further, even though land rental markets do not permit perfect adjustment to the desired size of operation (Skoufias 1995), the transaction costs incurred are smaller than in land sales markets. There has, however, been long standing con cern about the scope for incentive issues to lead to efficiency losses in rental mar kets. Under a land rental arrangement, tenants have few incentives to undertake long-term investments unless they receive the value of the investment back during the rental term or at its end (in the form of compensation) Moreover, with share tenancy, that is, in situations where the landlord receives a share of the harvest as rent, tenants receive only a fraction of their marginal product It is therefore diffi cult to motivate tenants to work hard enough, a phenomenon that is known as "Marshallian inefficiency." Share tenancy arrangements are still more efficient than wage labor, however. They may be an "optimal choice," given the constraints faced where markets for credit and insurance are incomplete For risk averse tenants (where risk aversion can arise out of the need to satisfy a minimum subsistence constraint), a share contract can provide insurance against fluctuations of output and income (Cheung 1969). For landlords, share tenancy insures against rent de fault by tenants whose wealth is insufficient to pay the rent or who, because they are credit constrained, underuse inputs (Shetty 1988) Thus the poorest tenants often receive wage contracts, while richer individuals (who have a lower risk of default) cultivate under share contracts with progressively increasing tenant shares fixed-rent contracts are limited to wealthy farmers (Laffont and Matoussi 1995; Lanjouw 1995; Shaban 1991).

The degree to which policymakers need to be concerned about preventing effi ciency losses from share tenancy contracts depends on the magnitude of this ineffi ciency. Although different methodologies produce widely diverging results, Shaban (1991) indicates that the losses may be relatively modest—about 16 percent Thus even if the government imposed regulations to replace the tenancy system with more efficient forms of production, the impact is likely to be modest. The case for govern ment intervention is further reduced because inefficiency decreases with monitoring

nd social control, that is, by embedding share contracts in long-term social or kinship relationships. In fact, Sadoulet, de Janvry, and Fukui (1997) find that share tenancy contracts between kin (but not between others) were not associated with any disincentive effects. The study of the efficiency implications of contracts should therefore be complemented by a focus on the contracting parties' opportunities outside their specific contract and on possible changes in the economic environment that might lead to the adoption of different types of contracts (Mookherjee 1997).

Implications for Land Sales and Rental Markets

When the 1975 land reform paper was written, policy advice focused on land sales markets to achieve efficiency-enhancing transfers of property and took a decidedly negative stance toward rental markets for land. If other markets are imperfect, however, as noted above, sales may be less efficient and less equitable than rentals. Even where attempts to liberalize markets for land sales are embedded in a well-sequenced program of integrated factor market development, land sales may not be the most important constraint on higher productivity, and their liberalization may not be an immediate priority.

Consider the effect of land transfers in rental markets. By renting out, landowners would not forgo possible benefits from credit access associated with land ownership (and could even advance the credit thus obtained to a credit-constrained tenant under an interlinked contract) and would at the same time benefit from any efficiency advantage of the tenant's family labor. Credit market imperfections tend to affect rental markets less than they do sales markets; moreover, rental markets are associated with lower transaction costs and generate positive externalities by facilitating the acquisition of agricultural knowledge by the tenant and adaptation to changing labor availability. Thus, rental markets may contribute more to efficiency than sales markets (Carter and Olinto 1996). Governments should therefore aim to create conditions conducive to the development of rental markets, rather than implicitly or explicitly restricting the scope for tenancy, as they have done in the past.

Land Reform

The World Bank's 1975 policy paper strongly supported redistributive land reform on equity and efficiency grounds, pointing to the success of Asian land redistribution and the Kenyan "million-acre scheme," which redistributed land from European settlers to African farmers. The practical difficulties associated with implementing land reform notwithstanding, the conceptual attractiveness of such a policy rests on three pillars.

First, in situations where credit and product markets are incomplete, access to land can make a significant contribution to food security, households' nutrition, well-being, and their ability to withstand shocks (Bardhan, Bowles and Gintis forthcoming). Evidence from China, where land was distributed largely independently of economic status, suggests that even though access to land insures household income only moderately against shocks, it provides almost complete insurance against malnutrition (Burgess 1997). Second, landownership affects economic growth and poverty reduction through credit-financed investment. The underlying idea is that the lack of collateral precludes landless individuals from making investments (in education, livestock, wells, and so on) that would require credit, even though the investments would profit both the individual and society (Eckstein and Zilcha 1994; Galor and Zeira 1993). Poor people who do not have access to assets might remain impoverished not because they are unproductive or lack skills but because they never get the opportunity to utilize their innate ability (Fafchamps and Pender 1997; Jalan and Ravallion 1997). And finally, several studies have argued that a more egalitarian distribution of assets (not necessarily land) would improve political stability. Because this issue does not relate directly to land issues, we refer the reader to the literature on this topic.[3]

The ease of actually implementing land reform has varied considerably between "landlord estates," which had been cultivated by tenants, and "haciendas," whose tenants received a small plot of their own in return for working on the landlord's farm. In landlord estates, all that is required is a reassignment of property rights; land reform is generally easy to implement, and stable systems of production emerge. Since the end of World War II, landlord estates in Bolivia, large areas of China, Ethiopia, eastern India, Iran, Japan, the Republic of Korea, and Taiwan, China, have been transferred to tenants in the course of successful land reforms. The productivity gains associated with these land transfers were modest in cases where security of tenure had already been high, where cash rent (rather than share rent) contracts had prevailed before the reform, and where landlords had provided tenants with market access (and no subsidization was available). Both welfare and productivity increased where investment opportunities were available (Callison 1983; King 1977; Koo 1968), where land ownership enabled the new owners to access markets for credit and insurance that had previously been beyond their reach (Dorner and Thiesenhusen 1990), and where new technology could be readily adopted (Otsuka, Chuma, and Hayami 1992).

By contrast, land reform in hacienda systems has been very difficult, and the "game of Latin American land reform" has been declared lost (de Janvry and Sadoulet 1989). In the great majority of these systems, large landowners responded to the threat of land reform by evicting all hired workers or tenants who could have claimed ownership under a reform program. The landlords either switched to livestock production and ranching or—aided by significant credit subsidies—shifted to highly mechanized cultivation (Binswanger, Deininger, and Feder 1995). As a result, programs of

distributive land reform reached far fewer people than intended and were often companied by a decline in tenant welfare that may have outweighed the benefits of e programs. Several factors account for this lack of success.

First, if land is transferred from large to small farmers through government prorams, the ability of the latter to make economically productive use of this asset is ontingent on a change in the pattern of production, subdivision of the farm, and onstruction of complementary infrastructure. Second, because the main productivy advantage of land reform is linked to the increased incentives of owner-operators, t is important not only to avoid collective forms of production but also to ensure hat owners operate their own farms. Third, beneficiaries are unaccustomed to makng independent entrepreneurial decisions, an ability that is particularly important o make individual family farming a success. In many cases in which the farms acquired under a land reform program were not farmed at full capacity, the lack of unds for pastures, fencing, and so on or for startup capital was often the reason for he lack of success. Similarly, programs that were limited to transferring land to xisting workers without providing those workers with complementary investment, raining, technical assistance, and resources were generally associated with very limted equity and efficiency benefits.

Without access to credit markets, land reform beneficiaries may well be worse ff than they had been when the landlord provided them with inputs and possibly even credit for consumption smoothing (Guinnane and Miller 1997). Restricted access to credit together with insecure property rights led beneficiaries f land reform in Nicaragua (Jonakin 1996) and the Philippines to sell off their ew holdings—often at prices well below the productive value of the land. The ey to avoiding such an outcome is the ability to access output and financial markets (Brooks and Lerman 1994). Arrangements where financial intermediares provide input credit and help with marketing of the farm produce have in ome cases helped beneficiaries overcome the obstacles posed by market imperfections (Deininger 1999).

Implications for Policy: Communal Tenure Systems

Communal tenure systems are dominant in most countries of Africa, in China, in indigenous areas in Latin America, and in Mexico. When the community rather than the individual owns the land, whatever market exchanges (sale or rental) exist are normally limited to the community. Individuals have very secure and normally inheritable rights to land even after a period of absence, but they do not have permanent property rights to a specific plot, a limitation that may reduce investment incentives. In some cases, communal systems also permit periodic redistribution of land by the village chief to accommodate population growth.

In the past, communal tenure arrangements were often considered economically inferior and equivalent to collective production. The establishment of freehold title and the subdivision of the commons were proposed to prevent the efficiency losses that were assumed to be associated with communal ownership. More intensive study of communal tenure systems in a broader framework and the recognition that these systems perform multiple functions has led to a reassessment of these recommendations, however.

On the one hand, the efficiency losses associated with communal tenure systems may be more modest than generally assumed, for a number of reasons. First, arable land (in contrast to pasture, forest, or fishing grounds) is, in most communal systems, cultivated by individuals who enjoy inheritable rights, which means that the static (and maybe even dynamic) efficiency losses possibly associated with communal tenure may be quite limited. Second, communal resource ownership is often maintained because it either provides public goods or takes advantage of synergies that would be difficult to provide under individual cultivation, including risk reduction through diversification (McCloskey 1991; Nugent and Sanchez 1993), economies of scale to help with seasonal labor bottlenecks (Mearns 1996), and investment in community-level infrastructure (Boserup 1965; Dong 1996). Third, when population density is low and the payoffs from land-related investments are limited, the investment disincentives associated with communal tenure are likely to be of little consequence because people do not tend to invest under either system of tenure. With arable land becoming increasingly scarce, many communal tenure systems either recognize a user's property rights if the land has been improved or compensate the user for improvements when the land is redistributed, thus attenuating tenure-related investment disincentives (Sjaastad and Bromley 1997). Finally, although communal systems prohibit land transactions with outsiders, rentals and often even sales—within the community (and possibly beyond) are normally allowed, providing scope for efficiency-enhancing transfers.

On the other hand, in environments with low population density and limited access to infrastructure and markets, the costs of delimiting and enforcing boundaries for individual plots are high, so the economic benefits of formal titling may not offset the expenses involved. Indeed, in several African countries, titles that were generated at high cost have lost their value as landowners failed to keep them updated. Thus, in cases where there is no clear demand for demarcation of individual plots, communal titles that are administered internally in a transparent fashion could provide tenure security at a fraction of the cost of individual titles (Bruce and Migot-Adholla 1994; Heath 1992). Communal titles also might provide a more effective safety net to ensure against risks and substitute for more costly redistributive mechanisms (Burgess 1997). Thus, instead of reinforcing an often artificial dichotomy between private and communal rights or trying to privatize land rights to "modernize" land tenure in an environment where few of the conditions for such modernization

tion are present, policymakers should focus on ways to increase secure property rights within given constraints.

More secure land rights may be highly valued by cultivators even under conditions of relatively low population density. For example, in Zambia (with a population density of 12 people per square kilometer and where 75 percent of the land is suitable for farming), almost 50 percent of farmers feel their land tenure is insecure and would be willing to pay something (US$40, on average) for land titles (Deininger and others 1998). Disputes, efficiency losses arising from limiting transfers and barring certain groups from land rights, investment disincentives, and land grabbing in anticipation of future appreciation are all indicators that existing land rights are inadequate. Clarification and formalization of informal property rights in a process that increases the accountability of local leaders, establishes a transparent and implementable legal basis, and provides for adjudication of boundary disputes across communities must precede any effort to award formal titles. Adopting a flexible institutional structure that gives communities freedom of choice in accomplishing these goals is therefore of great importance. The draft land policy adopted by Zimbabwe provides a good example in this regard (Zimbabwe 1998).

In countries where land ownership has traditionally been vested in the state, policymakers are concerned that a shift to individual land ownership is likely to lead to an undesirable reconcentration of land ownership. Experience suggests that this concern can be accommodated without forgoing major productivity benefits by giving producers long term tradable leases rather than full ownership rights. For example, the household responsibility system in China (which gave 15-year lease rights and at the same time made individuals residual claimants to output) has led to tremendous increases in output and productivity. To increase investment incentives, the government has decided to replace the 15-year leases with 30-year contracts (Prosterman, Schwarzwalder, and Ping 1998). Because the degree to which earlier leases were honored varied greatly from village to village, inferences can be made regarding the impact of tenure security: studies find that more secure tenure did increase the level of investment (Jacoby, Li, and Rozelle 1998).

Establishing Formal Tenure Security

Land registration and titling have long been viewed as the main instruments for increasing tenure security, empowering a flourishing land market, and facilitating the use of land as collateral in credit markets. Although numerous studies have confirmed the positive impact of titling where the conditions are right, experience with World Bank projects has also demonstrated that titling is not a panacea for achieving a wide variety of divergent goals at the same time. The objective—whether it is to improve credit access, increase tenure security, or activate land markets—must be

clear. In addition, the ways in which individuals gain access to land before titling whether through collective, communal, or informal means, as well as the broader trajectory of economic development, will affect the costs and benefits of specific titling instruments, their incidence across population groups, and the scope for public intervention.

Conceptual and Implementation Issues

Improved credit access has repeatedly been shown to be one of the major benefits from formal title. Thus, titling will confer the highest benefits where informal land transactions are common, a formal credit market that permits use of title as collateral exists, and profitable investment opportunities are available.[1] Title is unlikely to increase the banks' willingness to lend to the rural sector where, for cultural or economic reasons, land cannot be repossessed or where land sales and mortgages are restricted (Atwood 1990; Ensminger 1997).

Even if the above preconditions are satisfied, the effect of titling may vary across groups of producers, an issue that is of particular relevance if the initial distribution of land endowments is unequal. If the transaction costs associated with lending to specific groups of producers exceed the benefits they can derive from the use of credit, title would not be expected to increase credit access. In such cases, the title might make it easier for large producers to access credit but would not make small landowners creditworthy, a situation that would deepen preexisting inequalities. To prevent this and help titling contribute to broad-based growth, concurrent measures to improve access to credit markets, and possibly a differentiated scheme of recovering the costs of establishing title, will be necessary.

If a case can be made for formal titling, it must be systematic and areawide to take advantage of economies of scale in measurement, adjudication, and conflict resolution. Similarly, complementary infrastructure (such as programs to ensure access to credit markets) can be provided more easily and cost-effectively under an area based program. To achieve equity, titling needs to be combined with a mechanism for resolving disputes on the spot and an information campaign explaining the legal background, the titling process used, the rights of different parties, the rules of evidence, and the benefits of the appeal process. If a decision is made to title on demand, the status of individual plots will still have to be investigated on a case by case basis, and any reduction in the transaction costs associated with titling will thus be minimal. Moreover, titling on demand has often had disastrous consequences for the poor because individuals with good political connections can often bypass the land rights of indigenous people, women, or other vulnerable groups (Bruce 1988; Platteau 1996).

The titling process requires a clear legal basis and a streamlined institutional infrastructure that is capable of administering the process efficiently. Numerous World

Bank projects have either underestimated the complexity of the technical issues involved in titling or assumed that titling could be initiated even if agreement over complex policy issues had not been reached. Many countries have a plethora of institutions, programs, and projects—often with overlapping competencies and responsibilities, contradictory approaches, and high resource requirements—that make it impossible to administer a titling program effectively or to instill confidence in the validity of the titles issued.

The absence of clear property rights increases the costs of land transactions and may drive them into the informal sector, but empirical evidence on the magnitude of this effect is limited, and government regulation of rental and sales markets appears to have been quantitatively more important. For example, in many Eastern European countries, land rental and even sales transactions emerged long before individuals were able to obtain formal land title. By contrast, the threat of expropriation of rented lands in Colombia and Mexico appears to have deterred land rental transactions even with a formal title. Evidence from Mexico suggests that formal individual title is not always necessary to facilitate operation of rental markets. The codification of property rights through proper procedures significantly reduced the transaction costs and increased the amount of land rentals in the market (Olinto, Davis, and Deininger 1999).

Examples

In the aftermath of the 1915 revolution in Mexico, about half of the national land area was granted to communities (*ejidos*) under communal title. Well-intended restrictions to prevent *ejido* land from falling into the hands of the wealthy proved to be highly inefficient. Although farmers invented ingenious ways to circumvent these restrictions (Heath 1992), commercial credit was difficult to obtain, and the transaction costs imposed by the various restrictions were high, involving, among other things, the threat of loss of land. In areas where nonagricultural opportunities had increased and farmers engaged in seasonal migration, communal tenure became increasingly dysfunctional.

This issue was addressed in 1992 by legislation that lifted the restrictions on transfers of land, subject only to an upper limit, and allowed *ejidos* to decide on the admission of members and the tenure regime under which they would operate. They can opt for communal tenure, contribute part of the common lands to a corporation or to a joint venture with outsiders, parcel all or part of the land out to members under freehold title, or even convert the whole *ejido* from communal to freehold tenure. Contrary to some fears, the law has not led to a widespread sell-off and pauperization of the majority of *ejido* members. Instead, allowing communities for which existing regulations had increasingly become a constraint the option to shift to a different tenure regime increased the owners' flexibility without giving up the core principles (and the advantages, such as the insurance function provided by joint land ownership) associated with communal types of tenure (de Janvry, Gordillo, and Sadoulet 1997).

The second example involves a reorientation of Bolivia's legal framework and overall institutional environment. Bolivia had a long history of arbitrary allocation of frontier lands to influential individuals. Corruption within the responsible agencies resulted in double titling and conflicting rights, which created considerable tenure insecurity. Titling was highly arbitrary; the process could take more than [illegible] years. The need to develop a broad consensus on the development of a more conducive land policy framework was time consuming, however, and required political support at the highest level, including presidential intervention (Munoz and Lavadenz 1997). Two areas of the resulting legislation are of particular interest.

Recognizing the possibility of multiple claims on land, the law established three classes of property rights: private homesteads, communal lands, and land under private freehold title. Homesteads are family residences up to half a hectare in size. As long as there are no conflicting claims, families can receive title to their homesteads merely by showing proof of possession and occupancy in a relatively simple and quick process. Title to communal lands (those that have been continuously used by a community or indigenous group) will be awarded if the community acquires legal status and specifies the by-laws under which it intends to govern internal matters. Private agricultural properties under freehold can be established on the remainder of the lands. In addition, public lands can be allocated free of cost to indigenous communities and landless peasants or, if there is no demand from these groups, sold at market prices through a competitive auction. The legislation declares all land titles that have been acquired illegally to be null and void.

Conflicting claims are adjudicated, land rights are regularized, and a legal cadastre is established by a newly founded institution. Areas to be reformed first are selected on the basis of existing land conflicts and demand for regularization. A desk investigation based on existing registry information and aerial photography is followed by a field investigation that involves all the claimants. The results of this investigation are then cross-checked with the land records and published in the communities to elicit public comments. Once complaints and objections have been attended to, the results are posted publicly for two months. If no further complaints are lodged, the rights are finally registered. Experience thus far suggests that all but a tiny minority of claims can actually be regularized using this process and that large landowners are happy to cede illegally acquired parts of their "property" in exchange for legally recognized title to the rest.

Improving the Functioning of Land Markets

Governments in many developing countries maintain regulations that restrict land use and transfers. In many cases these restrictions have been adopted to avoid [illegible] unequal concentration of landholdings under a distorted policy regime or to reduce [illegible]

[illegible]

the political and fiscal cost of implementing land reform. A review of these policies finds that they have rarely achieved their goals. We examine three main issues: restrictions on land ownership and use; restrictions on land sales and rentals; and interventions to improve the functioning of land markets.

Land Ownership and Use

Ceilings on land ownership have been imposed primarily to facilitate the breakup of large farms and the associated sales of land to small producers. Even where such measures have had a strong economic and social justification and where conditions for implementing them should have been favorable, ownership ceilings have had only a marginal impact. In India, for example, 35 years of ceiling laws have, in all except three states, transferred less than 1 percent of the agricultural area to the target group (Appu 1996). Ceiling laws have been expensive to enforce, have imposed costs on landowners who took measures to avoid them, and have generated corruption, tenure insecurity, and red tape.

Such ceilings might be justified as a temporary measure in situations such as Eastern Europe if there are large imperfections in markets for credit, inputs, and outputs and if (new) landowners are ill informed about the productive value of their endowment. In such a situation, measures to reduce the scope for rapid land accumulation by individuals with better market access or information might be justifiable—although a temporary sales moratorium may be a better way to achieve this than ownership ceilings.

Governments may also adopt zoning laws that classify certain land as either agricultural or nonagricultural. In rural areas, zoning land for agricultural use provides tax credits, exemption from assessments for urban services, and protection from nuisance suits and forecloses the option of selling the land as residential property. Zoning is justified if negative externalities need to be reduced by more than the cost of zoning enforcement (Brandao and Feder 1995), but the cost of enforcing zoning regulations that run counter to economic incentives should not be underestimated. Especially if the institutional infrastructure for enforcement is weak, zoning may lead to rent seeking and corruption that reduce the economic benefits to a point where they become negative.

Restrictions on Land Rentals and Sales

Sales restrictions have frequently been imposed on beneficiaries of land reform or on workers on formerly state owned land to prevent them from selling or mortgaging their land. These measures could be justified to prevent beneficiaries from taking undue advantage of a land reform program or as a temporary measure to prevent land sales based on a lack of information or in response to imperfections in product

and financial markets. Even temporary restrictions on land sales can be counterp ductive, however, because they prevent landholders from accessing credit when most needed. Hayami and Otsuka (1993) describe a situation in which farmers w forced to resort to inefficient arrangements such as usufruct mortgaging and w labor to gain access to credit. Precluding beneficiaries of land reform from renting selling their land is likely to prevent adjustments that reflect the settlers' abilities a could, if combined with restrictions on rentals, cause large tracts of land to underutilized. The goal of preventing small landowners from selling out in respo to temporary shocks would be better served by ensuring that they have access output and credit markets and technical assistance and by providing them with saf nets during disasters to avoid distress sales. A moratorium on land sales might justifiable to discourage speculative purchases, but alternatives such as limiting amount of land that can be allocated to one individual or requiring that the land cultivated before title is granted should be considered instead

Rental restrictions aimed at eliminating the efficiency losses that are assumed be associated with share tenancy are not justified. They should be eliminated beca rental markets are likely to acquire increased importance with economic develop ment (in most industrial countries, between 40 and 70 percent of all cultivated ag cultural areas is rented rather than owned). As noted above, the efficiency gains fro rental restrictions are likely to be modest even in the most desirable case and th danger of less favorable outcomes is high The historical root of most rental restr tions in developing countries is in tenancy reforms that sought to improve the st and welfare of the tenant farmer by imposing rent ceilings, awarding perman rights to tenanted land (subject to landowners' right to retention), and transfer land ownership to lands not claimed by landowners The inability to imple these reforms swiftly has negatively affected the functioning of rental markets most Latin American countries that tried to give tenants secure tenure landl thwarted the reforms by undertaking large-scale evictions or shifting to ranch highly mechanized cultivation, or the use of wage labor In India, tenancy refo meant to benefit the poor seem, in the aggregate, to have damaged them Althou the impact varies by state, tenant evictions associated with tenancy reform caused the rural poor to lose about 30 percent of the total cultivated area threatening landowners who lease with the loss of their land, the reforms have pletely undermined land access through rental markets (Appu 1996) Even tries such as Egypt and Uganda where tenancy reforms could be implemented ure to separate clearly the rights of landowners from those of tenants overlapping claims to the same piece of land, causing uncertainty and inhibiting investment. Landowners (who normally are precluded from raising rents) have incentive to invest, while tenants' rights cannot normally be used as collateral formal credit.

ing rental restrictions with a clear regulatory framework for land rental ould do much to improve agricultural productivity and the welfare of po-ants. Furthermore, in situations characterized by overlapping property rights from incomplete implementation of tenancy reforms, mechanisms allow-rties to come to a mutually agreeable solution—whereby one party buys ther or each party receives full property rights to part of the land—could stment and productivity.

rket Interventions

ernments now realize that the social benefits from better-functioning land rkets far outweigh the advantages of most of the restrictions that have his-een imposed on the operation of such markets. The key question, there-identify the most important impediments to better development of land nd to sequence their removal in a way that does not jeopardize the poor. cussions indicated that the key issues are to enhance investment by clarify-rty rights and establishing an institutional framework that guarantees the these rights, to increase efficiency by facilitating increased transferability sed rights in rental and possibly sales markets; and to improve the integra-nd and other financial markets. In addition, governments can consider a land tax and establishing land information systems.

tax that is enforced at the municipal level not only could provide an to large landowners to utilize their land more productively but could also important contribution to decentralization. On the one hand, a land tax the few cases of a lump sum tax where —using asset, rather than produc-es — the effective tax rate decreases as the income generated from the land thus encouraging more productive use of the resource. Several countries nly experimenting with a land tax, either using a flat tax rate as in Nica-basing land taxes on self assessed land values as in Chile (Bird 1974). s have proven very useful in a wide range of urban contexts in developing and if accompanied by appropriate institutions to help with account-implementation should be feasible in rural ones as well. Because the nd tax revenues in any given municipality is linked to land values, how-ntial land tax revenues will obviously be meager in poor and remote rural lities. Land taxes therefore cannot redistribute wealth from rich to poor roods, which means that local governments will need additional sources e if the interests of horizontal equity are to be served.

countries are also attempting to establish market information systems ld reduce transaction costs and improve the availability of information d prices and markets. These systems would help expand participation in

sales and rental markets and thereby improve the acceptance of land as collat[illegible] by financial institutions. Such information systems would also help in develop[illegible] fine-tuning, and evaluating the broader framework for land policy, particularly determining the degree to which distortions continue to apply, who exactly p[illegible] ticipates in these markets, and whether the interaction between land and cre[illegible] markets is efficient.

Redistributive Land Reform

Many of the impediments to a smooth functioning of land, labor, and product m[illegible] kets date from the colonial era; because such longstanding barriers maintain a high[illegible] unequal distribution of land, large tracts of productive land lie idle, while peas[illegible] have to eke out a living on marginal and often environmentally fragile lands [illegible] addition to reducing productivity, unequal land ownership is also linked to so[illegible] unrest and violence. But the practical difficulties of implementing a land refo[illegible] program and the ease with which the economic imperatives might be subordina[illegible] to political pressure are daunting obstacles.

In the past, instead of aiming to increase productivity and reduce poverty [illegible] reform often aimed at calming social unrest and allaying political pressures by [illegible] ant organizations (Horowitz 1993). Even where there was a genuine commitme[illegible] breaking the power of landed elites, agrarian reforms were generally designe[illegible] urban intellectuals with little idea of the realities of agricultural production [illegible] suspicion that small-scale cultivators could not farm on their own [illegible] crease productivity (Barraclough 1970). Moreover, the individuals who were [illegible] geted to benefit from these programs were often politically powerful and [illegible] connected rather than those who could make productive use of the land or who [illegible] the most deserving on poverty grounds.[1]

Furthermore, the continued existence of implicit and explicit distortions (for ex[illegible] ample, the use of land as a tax shelter) raised the costs of land reform by driving [illegible] prices above the capitalized value of the agricultural profits the land would pro[illegible] Such distortions also reduced the sustainability of land reform and, by encourag[illegible] beneficiaries to sell out to large farmers, contributed to a reconcentration of [illegible] ings. As noted earlier, attempts to impose legal restrictions often made matters wor[illegible] A recent census of Brazilian land reform settlements reported that only about [illegible] percent of land reform beneficiaries were actually tilling their land.

Finally, rather than improving the way land markets function and using these markets to complement government efforts to redistribute agricultural land, prev[illegible] ous programs often aimed to provide substitutes for these markets, resulting in c[illegible] plex regulations that stretched available administrative capacity (Lipton 19[illegible] tralized government bureaucracies --charged with providing technical assistan[illegible]

her support services to beneficiaries—proved to be corrupt, expensive, and ineffec-
ve in responding to beneficiary demands.

and Banks

view of these difficulties, land banks and frontier settlements were seen as alterna-
ve mechanisms to land reform. Land banks provide loan financing at commercial
tes for small farmers to acquire land, while frontier settlement, or colonization,
ms to transfer individuals from congested areas to remote areas where lack of infra-
ucture means that land is cheap. With hindsight it can be said that these alterna-
ve mechanisms were ineffective. Expecting beneficiaries to repay the full price of
nd has resulted in widespread default and nonrecoverable loans. Frontier settle-
ent is no longer seen as a way to equalize land distribution. In addition to high
inistrative costs and associated environmental hazards, it has reinforced, rather
an eliminated, unequal land ownership patterns in many countries (Thiesenhusen
91). Thus most land reforms have relied on expropriation and have been more
ccessful in creating bureaucratic behemoths and in colonizing frontiers than in
distributing land from large to small farmers, although redistributive land reform
is shown to have positive social returns.

ew Opportunities for Viable Reform

fall in land prices associated with macroeconomic reforms, along with the loss of
e privileges that had been conferred on large farms by discriminatory laws, trade
otection, and credit subsidies, provides an opportunity to address land reform that
is detrimental to the functioning of markets. Several countries (Brazil, Colom-
a, Guatemala, the Philippines, and South Africa) are experimenting with a new
mmunity based" model of land reform. In this instance, the government's role is
ited to providing groups of poor people with technical and financial assistance to
land in a way that is similar to demand-driven social investment funds. This
proach has a number of advantages. First, because there is an upper limit on the
unt of the grant, beneficiaries have an incentive to seek run-down, unproductive
ates. This approach also aims to replace the confrontational atmosphere that has
aracterized land reforms with a more collaborative attitude. In fact, because any-
ng that improves the buyer's productivity is likely to increase the land price, the
ler will, in a competitive market, have a strong incentive to help buyers improve
e quality of their product—for example, through technical advice and marketing
sistance. Second, in a clear departure from the traditional approach, the new model
ould stimulate, rather than undermine, land markets. Finally, by drawing on the
ivate sector, nongovernmental organizations, and the community to develop, fi-
ance, and administer projects, the approach promises to overcome some of the

informational imperfections that have plagued the implementation of land reform by government bureaucracies. This also would help to develop a menu of project options more attuned to the specific needs of different groups within the target population (such as female-headed households).

These programs are too new for their impact on productivity and on the poor to be assessed. But initial evaluations underscore both the potential and the importance of the incentive framework and close monitoring. In South Africa the lack of local government structures, the continued existence of the land subdivision act (which was repealed only recently), and a very centralized and bureaucratic process initially posed limits to private sector participation, the outreach of the program, and the economic viability of the projects. Based on this experience, efforts are under way to reduce the administrative requirements for "livelihood projects" that involve very limited amounts of subsidy, to strengthen incentives for beneficiaries' own contributions, and to decentralize implementation of the entire land reform program. The success of several "share-equity schemes," where beneficiaries form joint ventures with private investors (including former farm owners), together with evidence from land transactions in the market outside of the program (Graham and Lyne 199[illegible]) point toward considerable commercial potential for land reform

In Colombia evaluations show that the results of a community-based pilot program are clearly superior to those of previous programs and that formerly landless cultivators are able to establish highly productive agricultural operations (Forero 199[illegible]). The large size of the grant (70 percent of the land value), however, together with the legal requirement that it be used only to purchase land and not for complementary investments, reduces the economic and fiscal viability of the program. In addition to establishing small areas of perennials and vegetables, where productivity actually exceeds expectations, beneficiaries have purchased large tracts of relatively unproductive (pasture) land that often generates less revenue than is necessary to service interest on the debt (30 percent of the land value) incurred to purchase the land. Changing the program structure to avoid this problem and allowing each beneficiary family to purchase and invest in an area sufficiently large to fully occupy the family's labor (about 2 hectares) could greatly increase the economic return as well as reduce the fiscal cost.

In Brazil, where individual states sought to increase the pace of land reform, a pilot program to allow market-based acquisition of land by beneficiaries has had impressive results, accomplishing the land reform faster than expected. The new approach is now being implemented nationwide. Because of its decentralized nature, there is ample scope for innovative ways to ensure that the program is targeted to the poor, that it is economically viable, and that it provides incentives for repayment of the land credit, all issues that are of critical importance if the program is to be replicated on a broad scale (Buainain, Da Silveira, and Teófilo 1998; Navarro 1998).

Conclusion

Within the last two decades, considerable advances have been made in understanding the principles underlying land relations and in the way in which they might be affected by specific policy interventions aimed at growth and poverty reduction. At the same time, the number of countries where policymakers believe that the issues surrounding land relations must be addressed has expanded.

It is now recognized that formal title, under conditions of low population density, is not necessarily the most cost effective and desirable way to ensure secure tenure and facilitate land transfers. One alternative is to award property rights to communities, which then decide on the most suitable tenure arrangements. This system not only should reduce transaction costs but also should allow a more flexible evolution of the structure of property rights while at the same time restoring some of the traditional social functions of land through secondary common property uses. Evaluations of such approaches, which are in increasing demand all over Africa, would be highly desirable. Another option is to award long-term and transferable leases, which could increase investment and expand the scope for using the rental market to transfer land to more productive uses.

Experience shows that the undesirable outcomes that have been attributed to the operation of land markets were caused more by policy distortions and imperfections in other markets than by the operation of land markets per se. The fact that land sales are more affected than rentals by such factors suggests that the liberalization of rental markets should be a high priority. Indeed, the plethora of land market interventions has greatly reduced opportunities for the poor to rent land. A number of countries inherited a dualistic landownership distribution that is not conducive either to efficiency and investment or to equity and that has often been at the root of violent and protracted social struggle. After macroeconomic liberalization, some of these countries have started to implement a new model of community-based, market-driven land reform. Additional research is needed to determine whether such programs have affected land access, investment, productivity, and social indicators [illegible] as evidence. The results of that research not only will allow policymakers to make changes as individual programs evolve but also will provide lessons for countries that are struggling to make land policies more effective.

Notes

Klaus Deininger is an economist in the World Bank's Development Research Group; Hans Binswanger is director in the Rural Development and Environment Department in the Africa Region. The authors would like to thank Michael Carter, [illegible] Jarvis, Dina Umali-Deininger, Ruben Echeverria, [illegible] Feder, Gustavo Gordillo de Anda, Isabel Lavadenz, Shem Migot-Adholla, Pedro Olinto,

Elisabeth Sadoulet, and participants at a World Institute for Development Economics Research [illegible] workshop in Santiago and various World Bank seminars for detailed comments.

1. The breakdown in collective forms of production occurs because members do not receive [illegible] benefits of increased effort (the free-rider problem), because members' ability to benefit [illegible] collective's assets ends with termination of membership, thus diminishing members' investment [illegible] centives, and because there is an incentive to reduce the number of members, often [illegible] government subsidies to embark on a capital-intensive development path—implying that [illegible] generate much less employment than do small (or even large) farms.

2. Farm management and supervisory skills are of importance not only because farmers [illegible] management skills would operate larger farm units but also because they will generally [illegible] neration for their management comparable to what they could obtain in other sectors of the [illegible] This leads farm operators to substitute capital for labor as nonagricultural wage rates increase [illegible] and Peterson 1982). Such an increase in farm size over time does not necessarily indicate the [illegible] of increasing returns to scale

3. A positive relationship between asset distribution and growth is ascertained [illegible] Birdsall and Londoño (1997); Deininger and Olinto (1999), Deininger and Squire (1998) [illegible] Lederman, and Loayza (1998); and Rodrik (1998) Besley and Burgess (1998) [illegible] reform legislation.

4. Indeed, some studies have found that in cases where no formal credit markets existed [illegible] little impact on farm income or investment (Carter and Wiebe 1990 Migot-Adholla [illegible] 1991).

5. As in the case of Mexico, communities can decide to subdivide the communal lands [illegible] ute parcels to individual members under freehold title if they so wish subject to an upper [illegible] the holding of any individual in the group and adherence to proper processes in doing so

6. This is in line with evidence from a number of Eastern European countries where [illegible] constraints generally led to a relatively inefficient way of implementing land reform [illegible] physical restitution of plots rather than compensation of former owners through [illegible] payments

References

The word "processed" describes informally reproduced works that may not be [illegible] through library systems

Alston, L. J., G. D. Libecap, and B. Mueller 1999 *Titles, Conflict and Land Use: The [illegible] Property Rights and Land Reform on the Brazilian Amazon Frontier* Ann Arbor: University [illegible] gan Press

Alston, L. J., G. D. Libecap, and R. Schneider 1996 "The Determinants and Impact [illegible] Rights: Land Titles on the Brazilian Frontier" *Journal of Law, Economics [illegible]* 12(1):25–61

Appu, P. S. 1996. *Land Reforms in India: A Survey of Policy, Legislation and Implementation* [illegible] Delhi: Vikas Publishing House

Atwood, D. A. 1990. "Land Registration in Africa: The Impact on Agricultural Production" [illegible] *Development* 18:659–71

Baland, Jean-Marie, and Jean-Philippe Platteau 1996 *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* New York: Clarendon Press for the Food and Agriculture Organization.

Balcazar, Alvaro 1990. "Tamano de finca, dinamica tecnologica y rendimientos agricolas" [illegible] *Agropecuaria* 7(3):107–25

presented at the workshop on Reforma Agrária e Desenvolvimento Sustentável, Fortal[illegible] November (also at *http://www.dataterra.org.br/Documentos/documentos.htm*).

Burgess, Robin. 1997. "Land, Welfare, and Efficiency in Rural China." London School of [illegible] ics. Processed.

Cain, M. 1981. "Risk and Insurance: Perspectives on Fertility and Agrarian Change in Indi[illegible] Bangladesh." *Population and Development Review* 7:435-74.

Callison, C. S. 1983. *Land-to-the-Tiller in the Mekong Delta: Economic, Social, and Politi[illegible] Land Reform in Four Villages of South Vietnam.* Bloomington, Ind.: University Press of [illegible]

Carter, Michael R. 1984. "Identification of the Inverse Relationship between Farm Size and Pr[illegible] tivity: An Empirical Analysis of Peasant Agricultural Production." *Oxford Economic Paper[illegible]* 3[illegible] 45.

Carter, Michael R., and Dina Mesbah. 1993. "State-Mandated and Market Mediated Land R[illegible] in Latin America." In Michael Lipton and Jacques van der Gaag, eds., *Including the Poo[illegible]* P[illegible] ings of a symposium organized by the World Bank and the International Food Policy R[illegible] Institute. Washington, D.C.: World Bank.

Carter, Michael R., and Pedro Olinto. 1996. "Does Land Titling Activate a Productivity [illegible] Land Market? Econometric Evidence from Rural Paraguay." University of Wisconsin, M[illegible] Processed.

Carter, Michael R., and K. D. Wiebe. 1990. "Access to Capital and Its Impact on Agr[illegible] and Productivity in Kenya." *American Journal of Agricultural Economics* [illegible] 2 1146-50.

Carter, Michael R., and Eduardo Zegarra. 1995. "Reshaping Class Competitiveness and the [illegible] tory of Agrarian Growth with Well Sequenced Policy Reform." Agricultural Economics [illegible] per Series 379. University of Wisconsin, Madison. Processed.

Cheung, N. S. 1969. *The Theory of Share Tenancy.* Chicago: University of Chicago Press.

Deininger, Klaus. 1995. "Collective Agricultural Production: A Solution for Transition Eco[illegible] *World Development* 23(8):1317-34.

———. 1999. "Making Negotiated Land Reform Work: Initial Experience from Colombi[illegible] and South Africa." Policy Research Working Paper 2040. World Bank, Washington, [illegible] cessed.

Deininger, Klaus, and Pedro Olinto. 1999. "Is Broad Asset Distribution Good for Grow[illegible] Data Evidence from 60 Countries." World Bank, Development Research Group, W[illegible] D.C. Processed.

Deininger, Klaus, and Lyn Squire. 1998. "New Ways of Looking at Old Issues: Inequali[illegible] *Journal of Development Economics* 57(2):259-85.

Deininger, Klaus, Pedro Olinto, M. Wamulume, and D. Chiwele. 1998. "Agricultu[illegible] mance in a Post-Reform Environment: Implications for Second Generation Refor[illegible] World Bank, Washington, D.C. Processed.

de Janvry, Alain, and Elisabeth Sadoulet. 1989. "A Study in Resistance to Institution[illegible] Lost Game of Latin American Land Reform." *World Development* 1[illegible] 139-140.

de Janvry, Alain, Gustavo Gordillo, and Elisabeth Sadoulet. 199[illegible]. "Mexico's Second [illegible] form: Household and Community Responses, 1990-1994." University of Califor[illegible] ego, Center for U.S.-Mexican Studies, La Jolla, Calif. Processed.

De Meza, D., and J. R. Gould. 1992. "The Social Efficiency of Private Decisions to E[illegible] Rights." *Journal of Political Economy* 100:561-80.

Dercon, Stefan. 1996. "Risk, Crop Choice, and Savings: Evidence from Tanzania." *Econom[illegible] opment and Cultural Change* 44(3):485-513.

Dercon, Stefan, and Pramila Krishnan. 1996. "Income Portfolios in Rural Ethiopia and Tanzania: Choices and Constraints." *Journal of Development Studies* 32(6):850–75.

Dong, Xiao-Yuan. 1996. "Two-Tier Land Tenure Systems and Sustained Economic Growth in Post-1978 Rural China." *World Development* 24(5):915–28.

Dorner, P., and William C. Thiesenhusen. 1990. "Selected Land Reforms in East and Southeast Asia: Their Origins and Impacts." *Asian Pacific Economic Literature* 4:69–95.

Eckstein, Zvi, and Itzhak Zilcha. 1994. "The Effects of Compulsory Schooling on Growth Income Distribution and Welfare." *Journal of Public Economics* 54(July):339–59.

[illegible]nsminger, J. 1997. "Changing Property Rights: Reconciling Formal and Informal Rights to Land in Africa." In John N. Drobak and John V. C. Nye, eds., *The Frontiers of the New Institutional Economics*. San Diego, Calif.: Academic Press.

[illegible] M., and A. Kotwal. 1985a. "A Theory of Contractual Structure in Agriculture." *American Economic Review* 75:352–67.

———. 1985b. "A Theory of Two-Tier Labor Markets in Agrarian Economies." *American Economic Review* 75:162–77.

[illegible]champs, Marcel, and John Pender. 1997. "Precautionary Saving, Credit Constraints, and Irreversible Investment: Theory and Evidence from Semiarid India." *Journal of Business and Economic Statistics* 15(April):180–94.

[illegible] P., D. Lederman, and N. Loayza. 1998. "What Causes Violent Crime?" World Bank, Washington, D.C. Processed.

[illegible] Gershon. 1985. "The Relation between Farm Size and Farm Productivity: The Role of Family Labor, Supervision, and Credit Constraints." *Journal of Development Economics* 18:297–313.

[illegible] Gershon, and D. Feeny. 1991. "Land Tenure and Property Rights: Theory and Implications for Development Policy." *The World Bank Economic Review* 5(January):135–55.

[illegible] Gershon, T. Onchan, Y. Chalamwong, and C. Hangladoran. 1986. *Land Policies and Farm Productivity in Thailand*. Baltimore: Johns Hopkins University Press.

[illegible] Gershon, Lawrence J. Lau, Justin Y. Lin, and Luo Xiaoping. 1989. "Agricultural Credit and [illegible] Performance in China." *Journal of Comparative Economics* 13(4):508–26.

[illegible] Roberto. 1999. "Evaluacion de Proyectos Piloto de Reforma Agraria en Colombia." Informe [illegible] Santafe de Bogota. Processed.

[illegible] George B. 1994. "Does Supervision Matter? Some Hypothesis Tests Using Indian Farm-[illegible] Data." *Journal of Development Economics* 43:217–38.

[illegible] and Joseph Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Eco-[illegible]* 60:35–52.

[illegible] and Marcel Fafchamps. 1996. "Land Tenure and Allocative Efficiency in Niger." *[illegible] Journal of Agricultural Economics* 78:460–71.

[illegible] 1990. "Contract Farming and Outgrower Schemes in East and Southern Africa." *Journal [illegible] Agricultural Economics* 41:303–15.

[illegible] Andrew W., and Michael C. Lyne. 1999. "Land Redistribution in Kwazulu-Natal: An [illegible] of Farmland Transactions in 1997." University of Natal, Department of Agricultural [illegible], Pietermaritzburg. Processed.

[illegible] Timothy W., and Ronald I. Miller. 1997. "The Limits to Land Reform: The Land Acts in [illegible] 1870–1909." *Economic Development and Cultural Change* 45(3):591–612.

[illegible] Kisan, Sarah Williams, and Robert Romain. 1996. "Agricultural Credit Subsidies and Farm-[illegible] Values in Canada." *Canadian Journal of Agricultural Economics* 44(March):39–52.

Sadoulet, Elisabeth, Alain de Janvry and Seiichi Fukui. 1997. "The Meaning of Kinship [illegible] cropping Contracts." *American Journal of Agricultural Economics* 79:394–406.

Shaban, R. A. 1991. "Does the Land Tenancy Market Equalize Holdings?" Working Pap[illegible] sity of Pennsylvania, Department of Economics, Philadelphia, Pa. Processed.

Shetty, S. 1988. "Limited Liability, Wealth Differences, and Tenancy Contracts in Ag[illegible] nomics." *Journal of Development Economics* 29:1–22.

Sjaastad, Espen, and Daniel Bromley. 1997. "Indigenous Land Rights in Sub-Saharan A[illegible] priation, Security, and Investment Demand." *World Development* 25(4):549–62.

Skoufias, Emmanuel. 1995. "Household Resources, Transaction Costs, and Adjustme[illegible] Land Tenancy." *Land Economics* 71(February):42–56.

Thiesenhusen, William C. 1991. "Implications of the Rural Land Tenure System for [illegible] mental Debate: Three Scenarios." *Journal of Developing Areas* 26(1):1–24.

Udry, Christopher. 1996. "Efficiency and Market Structure: Testing for Profit Maxim[illegible] rican Agriculture." Northwestern University, Department of Economics, Evanston [illegible]

Yao, Yang. 1996. "Three Essays on the Implications of Imperfect Markets in Rural [illegible] diss. University of Wisconsin, Department of Agricultural Economics, Madison [illegible]

Zimbabwe, Ministry of Lands and Agriculture. 1998. "National Land Policy [illegible] Harare. Processed.

Zimmerman, Frederic J., and Michael Carter. 1999. "A Dynamic Option Value [illegible] Change: Marketable Property Rights in the Sahel." *American Journal of Agr*[illegible] 81(May):467–79.

Climate Change, Agriculture, and Developing Countries: Does Adaptation Matter?

Robert Mendelsohn • Ariel Dinar

ause most developing countries depend heavily on agriculture, the effects of global rming on productive croplands are likely to threaten both the welfare of the population d the economic development of the countries. Tropical regions in the developing world particularly vulnerable to potential damage from environmental changes because the oils that cover large areas of these regions already have made much of the land usable for agriculture.

Although agronomic simulation models predict that higher temperatures will reduce in yields as the cool wheat-growing areas get warmer, they have not examined the ability that farmers will adapt by making production decisions that are in their own interests. A recent set of models examines cross-sectional evidence from India and azil and finds that even though the agricultural sector is sensitive to climate, individual mers do take local climates into account, and their ability to do so will help mitigate impacts of global warming.

ientists have become more confident that greenhouse gases will lead to a rise in obal temperatures (Houghton and others 1996), developing countries have grown ough concerned about the economic impact of climate change on agriculture son and others 1996). Most of the empirical work to date has focused on the ustrial countries (Bruce, Lee, and Haites 1996; Reilly and others 1996), and al-ough experts have extrapolated the results of their findings worldwide (see, for ample, Fankhauser 1995, Tol 1995, or Pearce and others 1996), little research has used specifically on the developing nations.

To assess the likely effects of climate change, researchers have pursued three ap-oaches: agronomic models, agroeconomic models, and Ricardian models (which aw on the work by Ricardo showing that land values reflect a site's productivity).

The agronomic research that applies to developing countries focuses on the ability of farmland in less productive tropical climates (Rosenzweig and Parry Reilly and others 1996) and on the likelihood that warming will push ever farmland into this zone. Experts are further concerned that small-scale farm[illegible] have very little capital will not be able to pursue the new strategies that will [illegible]quired to adapt to the change in climate. Unfortunately, agronomic studies [illegible] examined the actual behavior of individual farmers. Farmer responses [illegible] hypothetical in these models, although the results demonstrate that adapt[illegible] significantly affect production outcomes (Rosenzweig and Parry 199[illegible]) [illegible] argue that only efficient adaptations should be included in forecasts of [illegible] pacts, claiming that farmers will adopt new methods only if the benefits [illegible] costs.

Agronomic models of climate sensitivity indicate that higher tempe[illegible] likely to be harmful in many developing countries where the climate is [illegible] water is inadequate, and temperatures are high (Rosenzweig and Parry 199[illegible] and others 1996). A further increase in temperatures will make many [illegible] areas less productive [illegible] and some completely unsuitable. In these model[illegible] made to examine the impacts of warming on all crops. Studies of [illegible] such as Rosenzweig and Parry's, have examined only grains and thus [illegible] account of crops that prefer tropical climates. Nor, as just noted, have [illegible] examined efficient adaptation. Thus such research may overestimate [illegible] from global warming. Their perspective has been limited to arbitrary [illegible] measures; even these, however, indicate that adaptation will have [illegible] outcomes.

Research suggests that climate change is not likely to have an adve[illegible] agriculture in the United States. The results come from two sources [illegible] and Ricardian models. Both approaches find that adaptation by farm[illegible] duce some of the damages from climate change. [illegible] examine farmers' alternatives and determine the most efficient [illegible] climate scenario. Kaiser and others (199[illegible]) and [illegible] and others [illegible] that farmers in midwestern regions of the United States will make de[illegible] ing crops, varieties, and farming practices to mitigate potential redu[illegible] The Ricardian models, which examined a cross section of farming [illegible] counties in the United States, found that the effects of higher temps [illegible] from mildly harmful to unequivocally beneficial (Mendelsohn, Nordh[illegible] 1994, 1996). Moreover, substantial evidence from laboratory and field [illegible] shows that elevated levels of carbon dioxide serve as a fertilizer and [illegible] growth and make plants more drought resistant. A doubling of the [illegible] bon dioxide in the atmosphere is predicted to raise crop yields by [illegible] percent (Reilly and others 1996). Although the exact magnitude of th[illegible] tilization is uncertain, the positive outcomes were almost universal [illegible]

ncorporates the assumptions of carbon fertilization and farmer adaptation, Adams nd others (1999) suggest that climate warming is likely to increase crop yields in the United States. These benefits, however, do not necessarily extend to the rest of the world because the climate in the United States is temperate and would remain temperate even with a little warming and because Americans rely on capital-intensive griculture that can adapt to a range of climates. It is not clear whether labor-ntensive agriculture has the same flexibility.

In an approach based on the Ricardian model, Dinar and others (1998) and Sanghi nd Mendelsohn (1999) use a cross sectional approach to examine the sensitivity of griculture in developing countries to changes in climate. By examining the actual erformance of farms across India and Brazil, these studies explore this question using the Ricardian model first developed for the United States (Mendelsohn, Nordhaus, and Shaw 1994). By regressing farm performance (land value or net income) on a set of environmental factors, traditional inputs (land and labor), and upport systems (infrastructure), it is possible to measure the contribution of each factor to the outcome and to detect the effects of long-term climate change on farm alues. Because farmers adjust to their local climates, the cross-sectional method automatically incorporates farmer adaptation. The results suggest that warming will be less damage than predicted by the agronomic models (which do not include adaptation). When Ricardian results are compared with agronomic models that do not [illegible] efficient adaptation, however, the results are consistent. It appears that the [illegible] model does a good job of including efficient adaptation in its predictions.

[illegible]ethodology

[illegible] techniques have been used to measure the impacts of global warming on agri[illegible] agronomic economic simulation, agroecological zone analysis, and Ricardian [illegible]sectional analysis.

[illegible] *[illegible] economic models.* Agronomic economic simulation uses a crop model [illegible] been calibrated from carefully controlled experiments in which the crops are [illegible] field or laboratory settings that simulate different climates and levels of [illegible] dioxide (Adams, Glyer, and McCarl 1989; Adams and others 1990, 1993, [illegible] [illegible]etching and others 1993, Kaiser and others 1993a, b; Rosenzweig and Parry [illegible] [illegible] and Parikh 1998b). Farming methods are not allowed to vary across [illegible] [illegible]ntal conditions so that all differences in outcomes can be assigned to the [illegible] that are being tested (temperature, precipitation, or carbon dioxide). The [illegible] do not include adaptation. The yields are then entered into economic models [illegible] predict aggregate crop outputs, prices, and net revenue. Because each crop re[illegible] [illegible] intensive experimentation, only the most important crops have been studied. [illegible] almost all of these studies have focused on grains. A notable exception is the

sure climate sensitivity in each country. Economic performance is measured us- farmland value in the United States and Brazil and annual net income in India.

vantages and Disadvantages of the Methodologies

archers using all three methods generally agree that the extent to which farmers pt to the new conditions can be very important. Agroeconomic and agroecological dels must explicitly model adaptation, however, in order to include it. The ana- must be able to determine which adaptations are economically desirable. In prac- such determinations are difficult to make, and so they have been done largely an ad hoc basis. The adaptation involves a change in agricultural practices in onse to a change in climatic conditions. It includes changes in management tices such as timing of sowing and harvesting, the intensification of inputs, and nges in the crop mix. Of course, adaptation assumes that farmers have access to [illegible] practices and technologies that are already practiced elsewhere.

ptation and Agronomic Studies

gronomic literature (which includes the agricultural components of the onomic economic approaches and the agroecological zone analyses) addresses tation by simulating changes in the growth parameters of various crops accord- o the latest scientific advances. This approach does not take into account eco- considerations and human capital limitations, both of which affect actual decisions. Therefore, it is hard to interpret the adaptation scenarios frequently [illegible] by agronomists. El-Shaer and others (1997) identify possible climate- ted [illegible] agronomists [illegible] adaptation strategies for Egyptian agriculture, but they do not provide quan- estimates of the changes in crop performance associated with these strategies. [illegible] and Rosenzweig (199[illegible]) identify several adaptation strategies for maize [illegible] including adjusted planting dates and the introduction of new maize [illegible] simulations for three sites suggest that earlier planting dates (10–30 days [illegible] the norm) increase yields by nearly 10 percent at all sites. The introduc- [illegible] varieties fully mitigated the negative impacts of climate change on yield [illegible] but only partially at the two southern sites. A combination of earlier [illegible]tes and new varieties completely offset the negative impact of climate [illegible] sites. Iglesias and Minguez (199[illegible]) evaluate several adaptation strategies [illegible] and maize in various climatological regions in Spain and find no reduc- [illegible] yields. The adaptation strategies they tested include combinations of new [illegible] changes in sowing dates, and double cropping, using short-cycle maize vari- [illegible] second crop along with lentils and a vetch–forage barley mixture. This [illegible] not only reduced the impact of increased temperatures on yields but also

permitted more intensive use of water and land. In Spain water efficiency im[illegible] by 1–10 percent in southern regions and 40–80 percent in northern regions [illegible] others (1994), who examine adaptation strategies in southern China, find th[illegible] rice cultivar increased yields at five out of seven sites. Changing the planting d[illegible] the currently used cultivars increased rice yields at the northern sites, but no[illegible] southern sites. Combining both changes—the cultivars and the planting d[illegible] significantly increased yields at six of the seven sites.

Schimmelpfennig and others (1996) and Lewandrowski and Schimmelp[illegible] (1999) review the literature that examines how farmers adapt to climate ch[illegible] the United States. For example, Kaiser and others (1993a, b) include ad[illegible] practices such as crop mix, crop varieties, sowing and harvesting dates, [illegible] saving technologies (tillage). Based on a comparison of nearby geographic [illegible] models (for example, Rosenzweig and others 1994) and farm-level models [illegible] adaptation reduced the negative impact of warming on crop yields by up [illegible] cent (Kaiser and others 1993a, b; Mount and Li 1994, Reilly 1994, 1995 [illegible] others 1996).

Pros and Cons of Agroecological Models

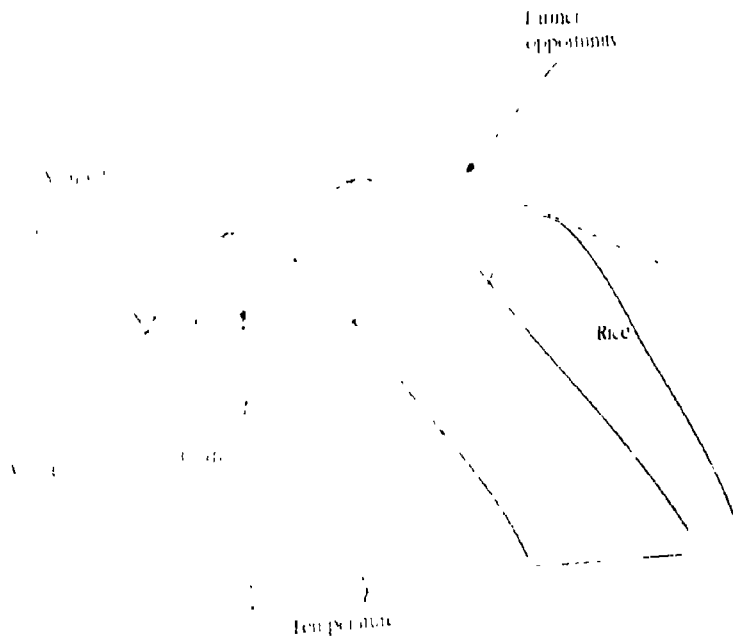
The biggest advantage associated with agroecological zones is that they [illegible] carefully studied and the geographic distribution of the zones in develop[illegible] tries has been published (FAO 1992). The current models using the zones [illegible] however, have many problems. The climate zones represent large temp[illegible] egories, so that subtle shifts within a zone have no effect but a small shi[illegible] zone to another has a dramatic consequence. The key measures of produ[illegible] not yet blended soils and climate together; the effect of each is comp[illegible] dently. Nor is it clear how tightly climate zones can predict which [illegible] grown or what their yields will be. The approach is subject to the same [illegible] the agroeconomic models in that researchers must explicitly account [illegible] Finally, the existing application of the method predicts large price [illegible] with small changes in aggregate supply, suggesting that there may be [illegible] the calibration of the underlying economic model (Darwin and others [illegible] win forthcoming). Although the technique has potential, the available [illegible] currently crude.

The Ricardian Approach

The most important advantage of the cross-sectional Ricardian approach [illegible] to incorporate the changes that farmers would make to tailor their ope[illegible] warmer climate. Because these adaptations benefit the farmer, there is [illegible] to expect that they will occur. One of the most crucial adaptation st[illegible]

hoice of crop. Depending on the effects of a warmer climate, a particular crop will e the optimal choice. Of the three grains shown in figure 1, for example, each is best uited for a specific temperature (and level of precipitation). If the temperature warms, owever, wheat yields will fall and net revenues will fall as well. If the farmer switches o corn, net revenues will rise. Optimal crop switching is an important component f measuring the agricultural impact of climate change.

One of the drawbacks of the Ricardian method is that the experiment is not carefully controlled across farms. Farms may vary for many reasons in addition to those corporated in the variables of interest. To control for this problem, the studies clude other important variables such as soil quality, market access, and solar radiaon. But it is rarely possible to get perfect measures of all these variables, and one nnot guarantee that all of these processes and interactions have been taken into count; some may not be measured at all. Paradoxically, this weakness is a strength the agronomic model, which relies on carefully controlled experiments and thus es not fall prey to this problem of extraneous variables.

gure 1. [illegible]

[illegible] calculations

Another valid criticism of the Ricardian approach is that it does not consider variations; all farms face the same prices. The models have consequently been [illegible] to assume that prices are constant, leading to a bias in the welfare calculations ([illegible] 1996). The cross-sectional Ricardian studies measure only the loss to producer ignoring the price change that would occur if supply changed, the approach [illegible] looks any loss in consumer surplus and consequently underestimates the dam[illegible] and overestimates the benefits.

Although it is easy to criticize these studies for assuming that prices are [illegible] stant, including price effects is difficult in any method. In most cases price [illegible] determined in a global market. A global model would be needed to predict [illegible] climate changes would affect crop yields. Unfortunately, we do not have [illegible] global crop models, so predicting what will happen to the global supply of crop as a result of a change in climate is difficult. Moreover, because the few gl[illegible] analyses completed to date predict that the range of warming in the 21st cen[illegible] should have only a small effect on aggregate supply, the bias from assuming prices are constant is likely to be small (Reilly, Hohmann, and Kane [illegible] and others 1996). For example, even if aggregate supply changed by [illegible] per[illegible] the bias from assuming constant prices would be less than [illegible] percent (Mendelsohn and Nordhaus 1996).

The Ricardian Approach Modified

The application of a cross-sectional approach to agriculture in developing [illegible] raises some additional difficulties that researchers addressed in the Brazilian [illegible] dian studies. Although many prices are constant throughout the sample, [illegible] not. Not only are these prices endogenous, but because they may not be [illegible] is difficult to control for their influence. For example, household members [illegible] large fraction of the agricultural labor pool in developing countries. No wages [illegible] paid to household members, nor are there data on the number of hours that [illegible] members work (Bennholdt-Thomsen 1982, Grepperud 199[illegible]). In Brazil and [illegible] therefore, researchers were forced to control for this factor using a dummy [illegible] that identified those farms that relied heavily on household labor. This dummy [illegible] able is difficult to interpret because it signifies unpaid labor, which implies [illegible] sign on net revenue, but the dummy also signifies a smaller and more [illegible] which implies a negative sign. Another input that is difficult to price is [illegible] In India, for example, there is an official price for a bullock [illegible] the purchase [illegible] other words. But bullocks also have to be fed and managed. Farms that [illegible] feed may find it cheaper to maintain a bullock than do farms that have to purchase animal feed. To try to control for the price of animal power, the Indian study [illegible] and others 1998) includes the number of bullocks per hectare as a control variable. Although this is an imperfect solution because the number of animals is endogenous

e researchers believe that it reduces the potential bias animal power may have on e climate coefficients.

Another issue on the input side is technology. Both India and Brazil have mounted rge and successful drives to enhance farming technology. These drives tended to be ncentrated on the more temperate farmlands in both countries. Farm technology nters were originally concentrated around São Paulo in Brazil and around the Ganges ver delta in India. Consequently, it was possible that the technology was facilitat-ng improvement in temperate climates, but not in tropical climates, making warm eas relatively less productive. McKinsey and Evenson (1998) examined this hy-othesis for India and found that technology had increased the output of farms over e last two decades, but because technological development had not specifically ddressed the problems of heat tolerance, the interaction between technology adop-on and climate appeared to be minimal.

Although technological adoption appears to have occurred in all climate zones in dia, new technologies can still affect climate sensitivity by giving farmers increased xibility (Antle 1995). The adoption of modern farming methods can free farmers om previous environmental constraints with new crop varieties, irrigation tech-logies, and chemical controls (Dinar and Zilberman 1991; Dinar, Campbell, and berman 1992). For example, figure 2 compares the climate senstitivity function the United States and India as it applies to Indian conditions. Both climate re-onse functions suggest that India is likely to suffer damages because initial tem-ratures are so high. The results using the Indian response function are more dam-ng, however, than the results from the American response function. It would pear that capital intensive agricultural systems are less sensitive to climate, perhaps cause they can control so many more inputs. Alternatively, the modern technolo-s may simply be more able to substitute purchased inputs for climate. Of course, e results also suggest that as developing countries get richer, their agricultural sec-r will become less sensitive to changes in climate.

esults of Studies

e agronomic results around the world vary, reflecting alternative methods, start-g conditions, and climate scenarios (Reilly and others 1996). Some of the studies luded some adaptation and carbon fertilization, whereas other studies ignored se factors. Some of the studies used old climate scenarios that tended to predict ge temperature changes, whereas other studies examined the more modest climate ulations now considered likely. Each of the studies was based on agronomic experi-ents in selected locations, and so they started from different initial conditions. ome of the studies were based on narrowly defined locations, whereas others ranged ver large territories. Although table 2 shows great variation across all regions, some

presented in table 4, for example, it allows a comparison of the impact of a change in climate on agriculture during the four seasons in Brazil and India. Net incomes in India decline precipitously with warmer winter, spring, and summer temperatures, whereas warmer fall temperatures increase net revenues. Land values in Brazil also decline with warmer summer and winter temperatures and rise with warmer falls. These results are similar to patterns found in the U.S. studies. The only seasonal exception is in Brazil, where warmer springs are beneficial. The harmful effects of warmer spring and summer temperatures in India are expected because the temperatures are quite hot already in India during this period. In Brazil, however, a warmer spring may simply extend the growing season. The effect of a warmer fall in all locations is expected to be beneficial as the higher temperatures help ripen and dry the harvest. The winter temperature effect is more controversial. Some agronomic models ignore winter temperatures because targeted crops are not growing at that time. Farm income may be very sensitive to winter temperatures, however, because cold temperatures help control pests. This can be important even if winter temperatures remain above freezing, as they do in most of India and Brazil. Net revenues are [illegible] sensitive to seasonal precipitation, but the effects are smaller and offsetting. [illegible] winters are beneficial, but wetter summer and springs are not. In India additional summer rains are not helpful because most of the country already enjoys a [illegible] monsoon during this period.

The cross-sectional studies [illegible] that the effect of climate change is not uniform [illegible] India. Even if the warming were the same throughout the country, some areas [illegible] heavily, most would be moderately damaged, and some areas would benefit [illegible]. Warming would damage the western coastal districts most heavily; districts [illegible] several eastern [illegible] along the coast would benefit. Interestingly, the desert [illegible] marginally dry areas [illegible] sensitive to warming; productivity in these [illegible] already so low that additional warming cannot harm them much further.

Policy Implications

[illegible] agronomic studies predict large agricultural losses in developing countries. [illegible] estimates appear to be too pessimistic for three reasons. First, the pessimistic [illegible] generally do not account for the powerful fertilizing effect of carbon dioxide. [illegible] these studies tend to underestimate the importance of efficient adaptation [illegible] mechanism to reduce damages. Third, almost all of the agronomic studies focus [illegible] which tend to prefer temperate climates, and do not include tropical and [illegible] crops. Taking these mitigating factors into account, climate change is [illegible] likely to reduce dramatically aggregate productivity in developing countries. [illegible] contrast, global warming is likely to increase productivity in industrial coun- [illegible] the temperate and polar regions. As these cooler regions become more pro-

ductive, the increased supply is likely to depress world prices, making farm[illegible] developing countries worse off. Although these price effects are likely to be [illegible] developing country agriculture will be relatively worse off.

The adaptation measured in the cross-sectional studies entails individual [illegible] but enlightened public policy could facilitate further changes. First, public [illegible] could help farmers adjust their cropping patterns and methods by monitoring [illegible] weather and providing better climate forecasts. Second, the government could [illegible] farmers on how to adjust to alternative climates. Third, the government could [illegible] in new technology by, say, funding research on heat-tolerant crops as an [illegible] introduce such crops into warmer climate zones.

Although scientific models provide important insights into the sensitivity [illegible] veloping country agriculture to changes in temperature, little is known about [illegible] regions, especially Africa and Oceania, and there is little information [illegible] warming will affect subsistence farmers. Agricultural GNP in developing [illegible] may not be severely damaged by warming because many developing countries, even those near the equator, have significant pockets of highly productive [illegible] ate farmlands. But poor farmers in marginal territories may be very vulnerable [illegible] ies are also vital to the future of other sectors of developing economies, such as energy, water, and coastal properties, all of which may be damaged by [illegible] Nonmarket factors such as changes in ecosystems, health, and aesthetics [illegible] considered. As we enter a world where climate is likely to change, it is [illegible] we learn as much as possible about the consequences of warming for [illegible] systems in order to prepare ourselves and to avoid the most serious imp[illegible]

Notes

Robert Mendelsohn is the Edwin Woolsey [illegible] Environmental Studies and Ariel Dinar is [illegible] ment of the World Bank. This research was funded [illegible]

1. Although we refer only to private adaptation [illegible] some management and other technologies [illegible] the result of public research and development.

2. Variables that were used in the Ricardian models [illegible] ing (not all nonclimate variables were used in both [illegible] monthly, seasonal) various measures of soil properties [illegible] capital variables, and technology level (measured as share of [illegible] suring infrastructure were hard to obtain at the level at which the analyses [illegible] were used in the form of population density and distance from market.

References

The word "processed" describes informally reproduced works that may not be [illegible] through library systems.

Houghton, J. T., L. Meira Filho, B. Callander, N. Harris, A. Kattenberg, and K. Maskell, eds. 19[illegible] *Climate Change 1995: The Science of Climate Change*. Cambridge, U.K.: Cambridge Univer[illegible] Press for the Intergovernmental Panel on Climate Change (IPCC).

Iglesias, Ana, and M. I. Minguez. 1997. "Modeling Crop-Climate Interactions in Spain: Vulner[illegible]ity and Adaptation of Different Agricultural Systems to Climate Change." *Mitigation and Adap[illegible]tion Strategies for Global Change* 1:273–88.

Jin, Z., D. Ge, H. Chen, and J. Fang. 1994. "Effects of Climate Change on Rice Production [illegible] Strategies for Adaptation in Southern China." In Cynthia Rosenzweig and Ana Iglesias, eds. [illegible]*plications of Climate Change for International Agriculture: Crop Modeling Study*. EPA 230 B [illegible] Washington, D.C.: U. S. Environmental Protection Agency.

Kaiser, H. M., S. J. Riha, D. S. Wilkes, and R. K. Sampath. 1993a. "Adaptation to Global [illegible] Change at the Farm Level." In H. Kaiser and T. Drennen, eds., *Agricultural Dimensions of [illegible] Climate Change*. Delray Beach, Fla.: St. Lucie Press.

Kaiser, H. M., S. J. Riha, D. S. Wilkes, D. G. Rossiter, and R. K. Sampath. 1993b. "A [illegible] Analysis of Economic and Agronomic Impacts of Gradual Warming." *American Journa[illegible] cultural Economics* 75:387–98.

Kapetanaki, G., and Cynthia Rosenzweig. 1997. "Impact of Climate Change on Maize [illegible] and Northern Greece: A Simulation Study with Ceres Maize." *Mitigation and Adaptation [illegible]gies for Global Change* 1:251–71.

Kumar, K. S. K., and Jyoti Parikh. 1998a. "Climate Change Impacts on Indian Agricultu[illegible] Ricardian Approach." In Dinar and others, eds.

———. 1998b. "Climate Change Impacts on Indian Agriculture: Results from a Crop Mo[illegible] Approach." In Dinar and others, eds.

Lewandrowski, Jan, and David Schimmelpfennig. 1999. "Economic Implications of Clima[illegible] for U.S. Agriculture: Assessing Recent Evidence." *Land Economics* [illegible]

McKinsey, James, and Robert E. Evenson. 1998. "Technology–Climate Interactions: Was [illegible] Revolution in India Climate Friendly?" In Dinar and others, eds.

Mendelsohn, Robert, and James E. Neumann, eds. 1999. *The Economic Impact of Climate [illegible] the Economy of the United States*. Cambridge, U.K.: Cambridge University Press.

Mendelsohn, Robert, and William D. Nordhaus. 1996. "The Impact of Global Warm[illegible] culture: Reply." *American Economic Review* 86:1312–15.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw. 1994. "The Impact of Glob[illegible] ing on Agriculture: A Ricardian Analysis." *American Economic Review* 84:753–[illegible]

———. 1996. "Climate Impacts on Aggregate Farm Values: Accounting for Adapt[illegible] *ture and Forest Meteorology* 80:55–67.

———. 1999. "The Impact of Climate Variation on U.S. Agriculture." In Mendelsohn [illegible] eds.

Mount, Tim, and Z. Li. 1994. "Estimating the Effects of Climate Change on Grain Yie[illegible] duction in the U.S." Cooperative Agreement, U.S. Department of Agriculture, [illegible] search Service, Washington, D.C. Processed.

Pearce, D., W. Cline, A. Achanta, Samuel Fankhauser, R. Pachauri, Richard Tol, and [illegible] 1996. "The Social Cost of Climate Change: Greenhouse Damage and the Benefits [illegible] Bruce, Lee, and Haites, eds.

Rao, N. G. P., G. R. K. Rao, and H. S. Acharya. 1989. "Yield Stability of Sorghum and M[illegible] Climates." In *Climate and Food Security*. Proceedings of the International Symposium [illegible] Variability and Food Security in Developing Countries, February 5–9. Sponsored [illegible]

[illegible]tional Rice Research Institute, Manila, and the American Association for the Advancement of Science, Washington, D.C.

Reilly, John. 1994. "Crops and Climate Change." *Nature* 367:118–19.

———. 1995. "Climate Change and Global Agriculture: Recent Findings and Issues." *American Journal of Agricultural Economics* 77:727–33.

Reilly, John, N. Hohmann, and Sally Kane. 1994. "Climate Change and Agricultural Trade: Who Benefits, Who Loses?" *Global Environmental Change* 4(1):24–36.

Reilly, John, Walter Baethgen, F. E. Chege, Siebe C. van de Geijn, Lin Ferda, Anna Iglesia, Gravin Kenny, David Patterson, Juta Rogasik, Reimund Rotter, Cynthia Rosenzweig, Wim Sombroek, and John Westbrook. 1996. "Agriculture in a Changing Climate: Impacts and Adaptations." In Watson and others, eds.

Rosenzweig, Cynthia, and M. L. Parry. 1994. "Potential Impact of Climate Change on World Food Supply." *Nature* 367:133–38.

Rosenzweig, Cynthia, B. Curry, J. T. Ritchie, J. W. Jones, T.-Y. Chou, R. Goldberg, and Ana Iglesias. [illegible] "The Effects of Potential Climate Change on Simulated Grain Crops in the United States." [illegible] Cynthia Rosenzweig and Ana Iglesias, eds., *Implications of Climate Change for International Agriculture: Crop Modeling Study.* [illegible] 230-B-94-003. Washington, D.C.: U.S. Environmental Protection Agency.

[illegible] Apurva. 1998. "Global Warming and Climate Sensitivity: Brazilian and Indian Agriculture." [illegible] diss. University of Chicago, Department of Economics, Chicago. Processed.

[illegible] Apurva, and Robert Mendelsohn. 1999. "The Impact of Global Warming on Brazilian and Indian Agriculture." Yale School of Forestry and Environmental Sciences, New Haven, Conn. Processed.

[illegible] Apurva, Robert Mendelsohn, and Ariel Dinar. 1998. "The Climate Sensitivity of Indian Agriculture." In Dinar and others, eds.

[illegible] David, [illegible] Lewandrowski, John Reilly, Marinos Tsigas, and Ian Parry. 1996. [illegible] Agricultural Adaptation to Climate Change. [illegible]-740. U.S. Department of Agriculture, Economic Research Service, Washington, D.C. Processed.

[illegible] Richard S. J. 1995. "The Damage Costs of Climate Change: Toward More Comprehensive [illegible]." *Environmental and Resource Economics* 5(4):353–74.

[illegible] Robert, M. Zinyowera, R. Moss, and D. Dokken, eds. 1996. *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific Technical Analyses.* Cambridge, U.K.: Cambridge University Press for the Intergovernmental Panel on Climate Change (IPCC).

What Does Climate Change Mean for Agriculture in Developing Countries? A Comment on Mendelsohn and Dinar

John Reilly

Mendelsohn and Dinar review much of the important work on the implications of climate change for agriculture, focusing particularly on developing countries. Their message is that efficient economic adaptation significantly reduces the estimated effects of climate change. Few dispute that some amount of adaptation is likely and its *potential* contribution to reducing the negative impacts of global warming is [illegible]. One such study (Darwin and others 1995), which analyzed the global impacts using an ecozone (land class) methodology, found that without adaptation, average cereal production yields fell roughly 20 to 30 percent in four different climate scenarios. Through various channels of adaptation (modifying crops and techniques on existing farmland, shifting crops to new land, and responding to changing market conditions) these losses were reversed, resulting in small increases in production worldwide (1 to 3 percent) even before considering the positive effects of carbon dioxide fertilization (table 1). Striking, however, are both the initial shock in cereal production in the study reported in table 1 and the range of impacts on yields (without adaptation) estimated by a variety of studies for different sites around the world (see table 2).

The Ricardian method reported by Mendelsohn and Dinar and the ecozone (land class) method of Darwin and others (1995) are similar in that they use cross-sectional data to estimate the adaptation response to climate change that occurs over time. Darwin and others (1995) use this evidence to estimate productivity shocks that are introduced into a general equilibrium model. As Mendelsohn and Dinar note, the Ricardian method is limited because it does not account for market effects, that is, the adjustment of prices reflecting market conditions. The result is thus strictly applicable only to a closed economy. Mendelsohn and Dinar note that this bias will be small if the global price effect is small, and they cite a study by Reilly, Hohmann, and Kane (1994) that shows small price effects in some scenarios. This single study is

change (in either direction) and hence no way to determine the direction or magni-tude of the bias. If adaptation proves to be as effective as Mendelsohn and Dinar, Darwin and others (1995) estimate, and if the CO_2 fertilization effect does increase yields by 10 to 15 percent, the prices of agricultural commodities may, in fact, *de-cline* sharply. Although a price decline would certainly be an economic benefit to consumers, agricultural exporting nations could sustain significant welfare losses. Reilly, Hohmann, and Kane (1994) make the point that exporting countries [illegible] the largest per capita losses (among cases with CO_2 fertilization and adaptation [illegible] they examine) under scenarios in which the world prices of agricultural commodi[illegible] fall. The point is that nobody has good estimates of the global impact. If the [illegible] to provide guidance for individual nations or regions, then results based on th[illegible] pothesis that the net global impact will be zero (or at least small) must be [illegible] with extreme caution.

The list of concerns about using evidence from cross-sectional data to estimate [illegible] impact of time-series phenomena is long. One problem is that of controlling [illegible] the other phenomena (either included in the estimated relationship but poor[illegible] sured or not included for lack of data) that might be affecting the estimated [illegible] ship between climate and agricultural production. Nordhaus (1996), who [illegible] gated the relationship between wages and climate to get at the direct value of [illegible] in people's everyday lives, used sophisticated econometric techniques to obt[illegible] ter estimates of the parameters. The study showed that the impact of global [illegible] on climate amenities could not be reliably determined. (As used here, climat[illegible] nity refers to the value people place on living in a warm and sunny climat[illegible] than a cold and snowy or hot and humid climate.) The relation between agri[illegible] productivity and climate in cross-sectional evidence would seem to be mu[illegible] ger, on the face of it, than the relation between wages and climate. Never[illegible] more robust measures of the reliability of the statistical estimates would [illegible]

A second major concern with cross-sectional evidence is that it represen[illegible] long-run equilibrium response. The Ricardian method and similar [illegible] approaches do not provide much information on how one gets from poi[illegible] rent climate and current production practices) to point B (new climat[illegible] production practices). Darwin and others (1995) provide a bit more insigh[illegible] channels of adaptation by dividing the response into three categories [illegible] occur on the farm, in the market, and in land use. Although these distin[illegible] somewhat artificial, they show that farmers are able to adjust even with[illegible] market response and without moving agricultural production to entire[illegible] Table 2 illustrates, however, that *without* adaptation, the impacts at individ[illegible] can be dramatic in both directions. Although tables 1 and 2 are difficult to compare directly, if one assumes that the overall picture presented in table 2 is roughly consis-tent with the "no adaptation" column in table 1, it appears that at finer geographic detail the response can be much more varied. In fact, many of the crop [illegible]

mates in table 2 were part of a study by Rosenzweig and Parry (1994), which, when aggregated to a global estimate, generated reductions in yields that were almost identical to those reached by Darwin and others (1995) in the case of "no adaptation."

In comparison with the reduced-form statistical approaches, agronomic models provide evidence on which technological solutions would increase yields (for instance, more fertilizer, changes in the planting date, new varieties of crops), but they do not offer any insight into whether farmers will actually choose these techniques or even whether these strategies would be economically beneficial responses. Reilly and Schimmelpfennig (forthcoming) point out that the techniques used by most studies contain hypotheses about whether adaptation will occur autonomously or not. Hence Mendelsohn and Dinar are concerned that crop response models introduce adaptation in an ad hoc manner, whereas cross-sectional evidence *assumes* agents will adapt to the changed climate even in a highly variable environment and will know which adaptations will work. Time series data can be misleading as well because they capture the response to unexpected weather events, whereas in the process of climate change agents may learn that some of these events are becoming more or less frequent and thus decide to adapt. If one assumes that dynamics do not matter, as implied by the use of cross-sectional evidence, then adaptation can and should be left to the market. If detection is expected to be difficult and agents need to learn the correct response, then the cross-sectional evidence shows the ultimate potential of adaptation. But public policy actions may be needed to realize this potential fully. If variabilities that slow the adaptive response are present, the costs may be greater than those estimated by cross-sectional methods unless or until the climate stops changing. Thus I believe it can be said only that the *potential* of adaptation is [illegible]

The growing literature reviewed by Mendelsohn and Dinar and presented briefly [illegible] raises at least three broad questions. First, how are these estimates to be used—[illegible] policy relevance are they? Second, how certain are researchers of these estimates? Third, given these estimates, what should be done now?

What Is the Policy Relevance of These Estimates?

[illegible] research agenda behind much of this climate change work is to develop estimates that clarify the damages associated with increased greenhouse gas emissions and the benefits of reducing emissions as proposed, for example, under the Kyoto Protocol of the Framework Convention on Climate Change (UNFCCC 1998). Integrated assessment efforts sometimes represent the problem as a generalized and dynamic cost-benefit analysis, where the benefits of the mitigation policy are the avoided damages to agriculture, coastlines, health, and other sectors (Nordhaus 1998). Most of the estimates focus on climate change associated with the equivalent of doubling

the pre-industrial levels of CO_2 in the atmosphere, with global average temperatu changes of 2.5° to 5.2° C.[1] The low end of this temperature range is not predicted occur until 2070; the high end is not predicted until well after 2100.

A push to foster adaptation through research on the likely effects of both clima change and adaptations to that change has been growing, for three reasons. First, may be economically sensible to spend something on adaptation and a bit less reducing greenhouse gas emissions. Second, if one despairs about reducing emi sions, given the costs and difficulties of reaching and enforcing a global agreemen adaptation may be the only defense. Finally, because inertia in earth and energ systems means that several decades of climate change are virtually inevitable, thos who are ill prepared to adapt (either to avoid losses or to take advantage of ne opportunities) may lose comparative advantage to those who are better prepared fact, work by Rosenzweig and Parry (1994), as reported in Reilly and others (199 shows the paradoxical result that cereal production in developing countries was low with adaptation than without. This decline occurred because the adaptation respon was stronger in the industrial countries. As a result, world prices were lower, agric tural comparative advantage shifted to the industrial countries, and developing co tries had less incentive to grow cereal crops. This finding does not mean that ada tion is a bad idea—if developing countries had not adapted at all, the shift w have been greater. It does indicate, however, the danger of basing results on pa equilibrium models or even on market or general equilibrium models of a si nation or region.

The general conclusion that adaptation (to the extent it is economically justi makes sense is tautological. But the value of the empirical work for identifying p ticular adaptation options is negligible or nonexistent. First, most of the work sumes that adaptation occurs without intervention from anyone. Researchers fig tively position themselves in low Earth orbit and observe that food continues produced and people continue to inhabit the land. The contrast between the in table 2 and those reviewed by Mendelsohn and Dinar (or between the fi last columns in table 1) suggests that something quite powerful must happ from estimated yield losses of 20 percent (or more) to the conclusion that eff minor or positive for the globe and for most countries. To the extent tha interested only in adding up the damages, perhaps one can assume that ev that needs to happen will happen. But part of what may need to happen is fo researchers to muck around on farms, in agribusiness, and in government a tural institutions to help point the direction.

Second, the time frame of 2070–2100 and beyond is irrelevant for decisions to about possible adaptation measures. Most of the capital in agriculture will be placed several times over in the next 70 years. It would be nonsense to optimize a system today for conditions far in the future and ignore the next three decades. It would be nonsense to optimize for conditions in 2070–2100 when most decision

n wait until 2069 or at least 2050, when much better forecasts will be available (if r no other reason than that the conditions in 2050 will already be known). Even here the lifetime of a project is long (for example, a large dam), almost any positive scount rate will make irrelevant to today's decision the question of whether there is iter in the river in 2100 or farmers who need it.[3]

Third, the level of uncertainty in these forecasts is unknown. For any particular untry, evidence and other simulations of doubled CO_2 effects suggest that pre-cted crop yields will vary, in either direction, by up to 100 percent of the nation's erage predicted yield under the same scenario. I discuss some of the reasons for this rge, subjective assessment of uncertainty later. If the assessment is reasonable, this vel of uncertainty poses significant challenges for the development of adaptation rategies. It is extremely dangerous to develop a strategy based on two or three sce-nios when so little is known about where these sit within the distribution of pos-le outcomes.

Fourth, these studies are insufficiently detailed or too incomplete—or both—to of much guidance. In work using these crop response model results and a fairly uled food trade model, Reilly, Hohmann, and Kane (1994) showed that in most intries the economic impact had as much or more to do with the effect of climate inge on world prices as with the impact of climate on agricultural yields within the country. In fact, net exporters of agricultural commodities generally benefited eco-mically from climate change if world prices rose (climate change was, on net, bad world production) even if they suffered yield losses. They suffered economic sses if world prices fell regardless of whether the climatic effects on agriculture in country were positive or negative. The situation for food-importing countries is reversed. The difficulty with the argument in Darwin and others (1995) suggest-g that global changes in prices may be small is that their study aggregates agricul-ral commodities to only three categories—grains, other crops, and livestock—and annot begin to investigate realistic changes in comparative advantage in the key port crops that are important for specific countries.

Some of these limitations affect the usefulness of these forecasts even for the global st-benefit calculus. It would be useful to have uncertainty bounds and to know more than just a few point estimates of impacts 70 to 100 years in the future. The limitations e fatal for adaptation actions other than the most general. It would be more useful to commend climate monitoring, more research, or better forecasts—but even for these, is unclear how much more money and effort should be spent. When researchers are ked to come up with robust strategies, the adaptation story is similar to the literature on reductions in emissions. In other words, researchers should look for adaptations that will improve resiliency to existing weather variability—so-called no-regrets adap-tations. Even such seemingly innocuous recommendations might go wrong. One might ell regret investing heavily in irrigation to reduce vulnerability to drought if climate change means that the river itself will dry up.

Are the Estimates of the Impact of Climate Change Valid

The body of work referenced here presents a somewhat negative result. Researc went looking for the impact of climate change, and even under the fairly extr scenarios of warming that might not occur until after 2100, they found little o effect. Logically then, if the problems, even in these extreme cases, are so slight warming between now and 2070 should have even smaller effects. One would conclude that it is unnecessary to reduce emissions or do anything else to adap climate change.

Are there errors in this logic? At issue is whether these are in fact "extreme" scen There are both socioeconomic and biophysical reasons why these scenarios are ex only in terms of average surface temperature change. Yet the evidence is that changes in temperature have little impact on agriculture production; extremes of perature, rainfall, and storm events are what cause negative agricultural outcom

On this subject the literature offers few strong conclusions. Will tropical s (hurricanes and cyclones) increase in number or intensity? The hydrologic will speed up; will that mean more intense rainfall and more frequent dro Will seasonal changes from cold to warm or from wet to dry become more va Will the El Nino-Southern Oscillation phases become more intense or remain in one phase for longer periods of time? Will monsoons and other patterns change their seasonal or geographic pattern? The lack of convinc dence forecasting these changes does not rule out any chance of their happen shift in rainfall patterns of 100 or so miles or by a month or two could le larger changes in precipitation in a particular region than is suggested by mated global average changes of [illegible] to 15 percent. Agricultural studies [illegible] imposed mean warming and precipitation changes from climate prediction climate of today without exploring the implications of the many dimensio mate that could change.

Are there catastrophes (low probability, high consequence events) th[illegible] set even the global mean estimates? The executive summaries of the rep Intergovernmental Panel on Climate Change (Bruce, Lee, and Haite[illegible] gues that the "consideration of risk aversion and application of the prec principle provide rationales for action beyond no regrets." The possibil strophic consequences occasionally enters discussions of climate chang ing of the West Antarctic ice sheet, a runaway greenhouse effect from th methane hydrates in permafrost or shallow coastal regions, and changes [illegible] conveyor belt are events that have been suggested, and in some cases [illegible] dismissed, as being highly improbable. If these are real possibilities, the[illegible] adaptation actions that governments could take to minimize the consequences strategies that would be helpful will not be understood if researchers persist [illegible] sidering only the center of the distribution.

think seriously about what should be recommended today. It is important to this question seriously, applying the model used in predicting hurricanes. If the ricane hits with no evacuation warning, the costs are high. But evacuating millio people if the hurricane turns away or if the forecasted point of landfall is imprec also disruptive and costly. Too many false alarms, and no one will believe the cast when it is right. Given the large degree of uncertainty in the estimates o effects of climate change on agriculture, researchers can only wish that their for were more precise. Additional work is needed to clarify whether adaptation w deed resolve any problems and to explore the full range of ways in which cli could change. A better assessment of the uncertainty involved in the forecas though not as helpful as a firm prediction) is more helpful than a firm pred that is wrong.

Notes

John Reilly is an associate director of the Joint Program on [illegible] at the Massachusetts Institute of Technology.

1. Equivalent doubled CO_2 refers to [illegible] spheric concentrations of CO_2 had doubled. Some of this [illegible] CH_4 (methane) and N_2O (nitrous oxide).

2. There are some well known issues with discounting [illegible] climate change (Lind and Schuler 1996). [illegible] ply judgments about intergenerational equity. [illegible] criterion that favors the poor should be biased [illegible] sumption to the present poorer generation. [illegible] problems with evaluating [illegible] at a different rate than other investments [illegible] largely involves normal investments [illegible] international commons problem of controlling [illegible] the compelling case that with uncertainty [illegible] at the lowest possible rate [illegible] be the "lowest possible discount rate." [illegible] then negative discount rates are possible. [illegible] context of the life of the investment. [illegible] with a lifetime of perhaps 50 to 100 years. [illegible] transportation infrastructure, power plants [illegible] declining discount rates for the distant future [illegible] would want to apply such a rate consistently [illegible] economy. And, under any circumstances, it is useful to know [illegible] of the investment rather than just a single [illegible]

References

The word "processed" describes informally reproduced works that may not be com[illegible] through library systems.